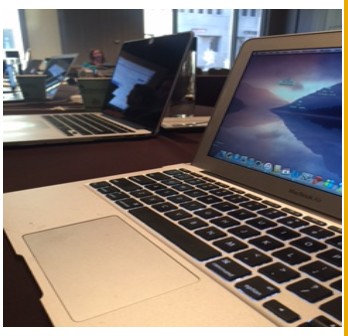
University of Victoria

# The Enduring Potential of Data

## An assessment of researcher data stewardship practices at the University of Victoria






*Prepared for:*

University of Victoria Libraries

*Prepared by:*

Dr. Jacqueline M. Quinless
Shahira A. Khair

January 2019

University of Victoria

# ACKNOWLEDGEMENT

We respectfully acknowledge the Lkwungen-speaking peoples on whose traditional territory the university stands and where the data for the study was gathered and this report was prepared. We also acknowledge with respect the Songhees, Esquimalt, and WSÁNEĆ peoples whose social-cultural and historical relationships with the land continue to this day.

The University of Victoria Libraries mirrors the University of Victoria's 2018 Strategic Framework values: excellence in all our endeavours; ethical and intellectual integrity; freedom of inquiry and freedom of speech; and equity, diversity, and inclusion. This report summarizes findings from a transdisciplinary study of research data management (RDM) practices at the University of Victoria and was prepared by Dr. Jacqueline Quinless, with support from Shahira Khair of the University of Victoria Libraries. This study was conducted from January 2017 to July 2018 in direct consultation with the Canadian RDM Survey Consortium, Council on Library and Information Resources (CLIR), Vice President Research, and the University of Victoria Libraries.

The study would not have been successful without the extraordinary efforts, professionalism, and support of many individuals. We are especially grateful to all of the researchers across the University of Victoria for sharing their time and perspectives about their research, our colleagues at CLIR in the United States, and our colleagues at the Canadian RDM Survey Consortium in Canada. Special thanks to all of the staff at the University of Victoria Libraries, especially Jonathan Bengtson (University Librarian) for his leadership and support, Lisa Goddard (Associate University Librarian) for her clear vision, and Lisa Petrachenko (Associate University Librarian) for her thoughtful ideas and teamwork approach. Thanks also to the following people for their research support and leading focus group discussions: Justin Harrison, Rebecca Raworth, Aditi Gupta, Tine Bebbington, Michael Lines, Ying Liu, In-In Po, and Tyne Ferreira.

# TABLE OF CONTENTS

# List of Figures

# List of Tables

# EXECUTIVE SUMMARY

## Project Background

The transition to conducting research in a digital environment requires the adoption of new practices and tools to ensure that research data are properly curated and managed, with the objectives of long-term security, accessibility, and reusability. While the adoption of digital methods has seen rapid expansion across most research disciplines, the development of knowledge, tools, and services to enable strong research data management (RDM) practices have generally lagged behind. In recent years, this gap has started to close, fuelled by the increased recognition of benefits to improved research transparency, productivity, and innovation that RDM enables. Emerging requirements for more responsible research mandated by granting agencies and scholarly publishers are also driving the need for better tools and services to support researchers with the management of their data.

In Canada, recent national policies governing federally funded research are shaping how both Canadian researchers and their home institutions will manage digital research data in the coming years. The *Tri-Agency's Statement of Principles on Digital Data Management* calls for excellence in digital data management and stewardship in agency-funded research (Government of Canada, 2016). More recently, the Tri-Agency have released a draft RDM policy that aims to set expectations and requirements for both researchers and their home institutions that administer grants and awards, relating to data management planning and long-term data storage (Government of Canada, 2018).

Against this policy backdrop, research libraries across Canada are exploring ways to work in collaboration with researchers to support their development of strong RDM practices. To help coordinate this work, the Canadian Association of Research Libraries launched the Portage network in 2015, which brings together members of research institutions, regional library consortia, and other key stakeholders, such as funding agencies and national infrastructure providers, to collaboratively address challenges and explore possible solutions for RDM in higher

education and research across Canada (Portage, 2018a). Through this network, the University of Victoria Libraries has joined a national survey consortium of Canadian university libraries focused on understanding data management practices within member academic communities by administering a common survey instrument to evaluate local RDM practices. By engaging in this work and this national network of expertise, UVic Libraries aims to inform its future services and infrastructure in order to meet demands for data management tools and support on campus.

## Purpose and Objectives

University of Victoria Libraries, with the support of the Vice President Research, has conducted an interdisciplinary research study, involving of all ranks of faculty members, post-doctoral fellows, and graduate students, in order to assess the current level of preparedness, as well as existing challenges and opportunities for improving RDM practices. Our findings will help to inform how the University of Victoria Libraries can facilitate data management activities on campus and support researchers across a range of disciplines. This report will also contribute to the ongoing national conversation on RDM practices in Canadian research institutions.

## Methodology

The mixed method approach taken by this study provides a rich source of quantitative and qualitative data that explores:

- how researchers manage and share their data
- differences in RDM practices across disciplines
- barriers to data sharing
- gaps in current infrastructure and services to support good data management

The data for this study were gathered using three methods:

1. A campus wide survey
2. In-depth interviews
3. Focus groups

The online survey was conducted from October 2017 to November 2017, yielding valuable insights from 418 participants from all major disciplines at the University of Victoria. Twenty-three in-depth interviews were held with graduate student and faculty researchers from various faculties from June 2017 to August 2018. In addition, three separate focus group sessions were held with twenty-one researchers and librarians at the University of Victoria Libraries, from June 2017 to September 2017.

## Key Findings

- Defining research data is not straightforward and there is considerable variation in vocabulary across disciplines. The term has different meanings depending on the faculty or discipline, the subject of study, the type of research being conducted, and the methodological framework being applied. Inconsistent definitions of what constitutes "data" can lead to broad and inconsistent interpretations of RDM activities and mandates.

- There is a clear demand for improved understanding of RDM practices among researchers, and a clear demand for related training and services. Researchers expressed interest in receiving assistance with preparing data management plans, and with documenting, securing, and archiving data.

- The majority of researchers indicated that they require guidance in order to complete a data management plan. As well, the majority of respondents were either unsure or did not believe their data had sufficient documentation to allow a person outside their own research group to understand and reuse it.

- Most respondents agreed there are benefits to sharing data. The top reasons indicated concerned benefits of collaborative scholarship, interdisciplinary research, research advancement, and open access to knowledge.

- Attitudes towards sharing vary across disciplinary groups. Respondents from the arts and sciences were more open to sharing and indicated fewer restrictions than their counterparts from medicine, law, business, and education. Respondents in these latter disciplines more frequently reported privacy, confidentiality, ethical, legal, or security reasons for not sharing data. As well, commercial and intellectual property concerns were identified as unique barriers for respondents from business and engineering disciplines.

- Researchers identified a variety of issues and concerns that prevent them from sharing their data, which include working with sensitive data, participant confidentiality, working with OCAP (Ownership, Control, Access and Possession) protocols in Indigenous communities, copyright law, and patent issues.

- While many researchers support data sharing in principle, they struggle with barriers related to costs, access to technology and storage, privacy concerns around sensitive data, culture challenges relating to the academic reward system, and concerns of improper use of shared data.

## Moving Forward

The data gathered through this study shed light on the current data management landscape at the University of Victoria, and identify existing knowledge gaps. The results will guide UVic Libraries in developing services and infrastructure that will help faculty to meet emerging funder mandates around RDM. The findings of this study can be used as a starting point for institutional action and will be shared with University of Victoria researchers, administrators, Libraries, and with the wider academic community. The aggregated results will also be shared with peer institutions nationally to inform services being developed by the Portage network, led by the Canadian Association of Research Libraries.

The following report recommendations are intended to support UVic researchers as they face new RDM mandates from funders and publishers, to facilitate sustainable RDM practices, and to foster data sharing among researchers.

## Recommendations for the University of Victoria

1. Provide clear guidance on funder requirements concerning RDM.
2. Increase researcher awareness of current institutional storage and backup options for working data.
3. Increase researcher awareness of Compute Canada default storage allocations, and assist researchers in gaining access to Compute Canada resources.
4. Provide discipline-specific guidance on standards for data description and formatting.
5. Identify RDM research champions at UVic to engage in RDM initiatives with the goal of expanding RDM capacity, expertise, and collaboration.
6. Increase the availability of sufficient, secure, easy to use storage solutions and RDM infrastructure to address current mandates and to meet future demand.
7. Continue to work nationally to advocate for increased funding for RDM infrastructure and expertise.

## Recommendations for the University of Victoria Libraries

1. Develop discipline-specific workshops and training materials to help graduate students understand the importance of RDM.
2. Offer direct project-based support to help research teams to develop strong data management plans.
3. Develop workshops to help researchers document their data for reuse in other contexts.
4. Offer direct support at various stages in the research life cycle to tackle specific curation issues at the beginning, midway, and after a research study.
5. Provide clear guidance on the distinction between active, archival, and repository storage and the role of each within the research data lifecycle.
6. Provide advice on repository options including discipline specific repositories, and repositories that are better suited for large data, or for particular file formats.

7. Offer guidance on journal policies and other scholarly communications requirements concerning RDM.
8. Promote the benefits of data sharing to university researchers, and help to remove some of the surmountable barriers to sharing.
9. Offer consultation on the retroactive sharing and curation of older data that may currently be at risk.
10. Work with faculty liaison librarians to determine their role in RDM support, and to better understand the specific needs of their departments and faculties.
11. Work in collaboration with RDM stakeholders across campus to improve communication channels, in order to effectively refer researchers to available supports and services.
12. Learn more about different community protocols, especially in the case of working with Indigenous data, to better address challenges and barriers to preservation.

# 1 CHAPTER ONE: INTRODUCTION

## 1.1 Background

Recent national policies governing federally funded research are shaping how both Canadian researchers and their home institutions manage digital research data. Since 2014, the Government of Canada's Action Plan on Open Government has aimed to maximize access to the results of federally funded research, in order to encourage greater collaboration and engagement with the scientific community, the private sector, and the public, with specific objectives for the open sharing of research data in standard accessible formats (Government of Canada, 2014). Meanwhile, the Tri-Agency's Statement of Principles on Digital Data Management, released in 2015, promotes excellence in digital data management practices and data stewardship in agency-funded research, focusing on the need for strategies to preserve and re-use research data (Government of Canada, 2016). More recently, the Tri-Agency released a draft RDM policy for consultation with the research community (Government of Canada, 2018). This policy aims to set expectations and requirements relating to data management planning and long-term data storage, affecting both researchers and their home institutions that administer grants and awards. The announcement of this draft policy, coinciding with the release of our report, will undoubtedly help frame the conversation about RDM at the University of Victoria going forwards.

Against this policy backdrop, research libraries across Canada are exploring ways to work in collaboration with researchers to support their development of strong RDM practices and enhance digital scholarship. To help coordinate this work, the Canadian Association of Research Libraries launched the Portage network in 2015, which brings together members of research institutions, regional library consortia, and other key stakeholders, such as funding agencies and national infrastructure providers, to collaboratively address challenges and explore possible solutions for RDM in higher education and research across Canada (Portage, 2018a).

## 1.2 Project Description

University of Victoria Libraries have joined a national consortium of university libraries who are committed to working together to understand and improve research data management (RDM) practices in Canada (Portage, 2018b). To support this goal, the University of Victoria Libraries have administered a campus-wide study that broadly surveys RDM practices. All ranks of faculty, as well as post-doctoral fellows and graduate students, from across disciplines were invited to participate in this study. Both quantitative and qualitative methods were applied in this investigation, with the intention of exploring:

- How researchers manage and share their data;
- Differences in RDM practices and needs across disciplinary groups; and
- How the University of Victoria Libraries can support researchers to enhance the quality of their digital data and RDM practices.

The data gathered through this study will help to expand the knowledge base of digital scholarship and data curation practices at the University of Victoria. They will also allow us to better understand researcher data curation needs and challenges at the University of Victoria, and help the library to develop services and infrastructure that will support faculty in meeting emerging funder mandates and publisher requirements concerning RDM. The findings of this study can be used as a starting point for institutional action and be shared among University of Victoria administration, libraries, research support services, and the wider academic community. The aggregated results will also be shared with peer institutions nationally to inform services like the Portage RDM network that is led by the Canadian Association of Research Libraries (Portage, 2018a).

## 1.3 University of Victoria Libraries

The library's long-standing and trusted role in enabling access to and preserving knowledge is enhanced by a focus on opening avenues to research, systems, and structures, and engaging actively with stakeholders. The library's strategic directions identify three core principles in the upcoming years as a primary focus: Open, Engaged, and Enduring.

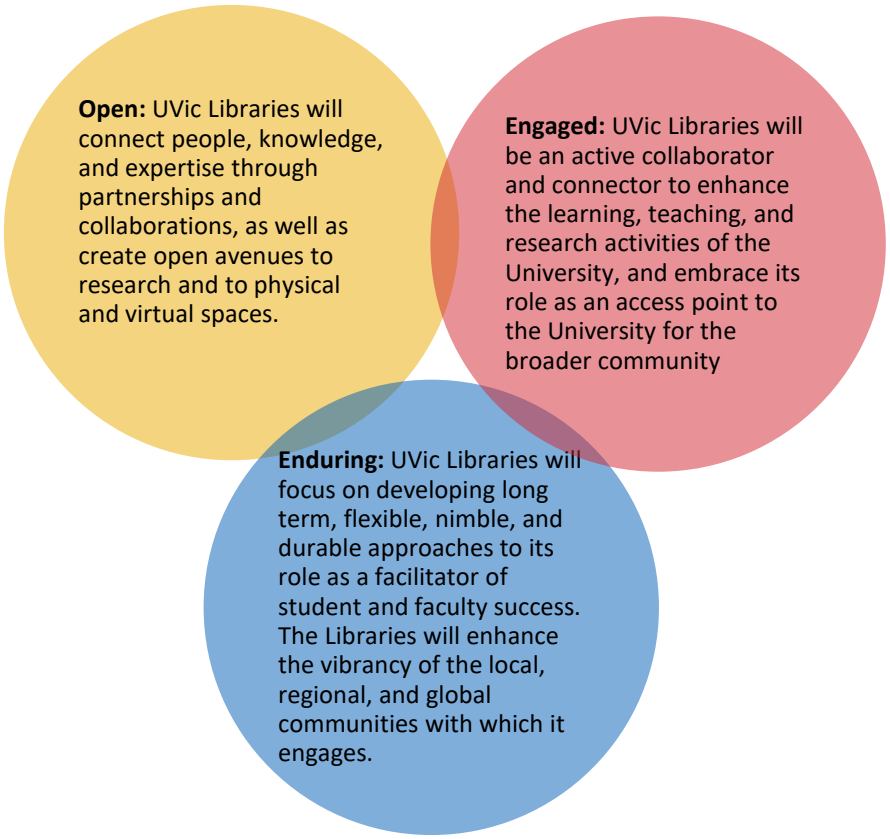



**Figure 1.** UVic Libraries' Strategic Directions 2018-2023[1]

To support the University of Victoria Libraries in meeting its strategic goals, as they relate to research data, expertise in data management, manipulation, and analysis should be developed in order to support the stewardship of research data unique to the University of Victoria. This report aims to inform the sustainable development of this capacity.

[1] https://www.uvic.ca/library/about/ul/strategic/index.php

## 1.4 Current RDM Landscape at the University of Victoria

Existing RDM services are currently distributed at the University of Victoria across several of different units:

**Office of Research Services**
https://www.uvic.ca/research/conduct/index.php

The Office of Research Services (ORS) assists faculty in securing and administering research grants, awards, and contracts, as well as meeting their regulatory responsibilities in support of research. ORS acts as the main Tri-Agency liaison body, and will have significant responsibility for ensuring that UVic researchers meet their RDM obligations under the pending Tri-Agency policy.

**Human Research Ethics Board**
https://www.uvic.ca/research/conduct/home/regapproval/humanethics/index.php

The Human Research Ethics Board (HREB) ensures that UVic research and research occurring in academic courses involving human participants or human biological materials meets the ethical standards required by Canadian universities and national regulatory bodies. HREB helps to identify sensitive or private data in research projects, and helps researchers to understand their obligations in the collection, management, sharing, and destruction of research data sets.

**University of Victoria Libraries**
http://libguides.uvic.ca/researchdata/home

UVic Libraries offer workshops and individual support for researchers who wish to use the nationally available DMP Assistant[2] to create data management plans for research projects. The libraries manage an institutional Dataverse repository[3] that is open to all UVic researchers. Subject liaison librarians and the data curation librarian develop and deliver RDM workshops for faculty and graduate students. UVic Libraries provide guidance and advice around all aspects of RDM, and help researchers to connect with data curation resources at UVic and beyond.

[2] https://assistant.portagenetwork.ca/en
[3] https://dataverse.scholarsportal.info/dataverse/uvic

### Research Computing (UVic Systems)

https://www.uvic.ca/systems/services/researchcomputing

Infrastructure Services hosts high-performance, high-availability mass storage for research purposes. Both online-disk and backup-tape storage systems are available for research computing users. This storage is accessible from the university's high-performance computing systems.

### Compute Canada

https://www.computecanada.ca/research-portal/accessing-resources/rapid-access-service/

Compute Canada's Rapid Access Service allows any Compute Canada user to access modest quantities of compute, storage, and cloud resources as soon as they have a Compute Canada account. UVic's high-performance computing specialist is available to help researchers access these resources and use them effectively.

### Department-level supports

Many departments offer some level of support for storing and backing up faculty research data, and may provide software tools for collecting, organizing, and analyzing research data. The level of available support varies from discipline to discipline, and department level IT policies are not always highly formalized.

### Research Computing Advisory Committee

The Research Computing Advisory Committee (RCAC) has representation from many of UVic's most data intensive disciplines, research centers, and research projects. The RCAC advises the university on needs related to research computing infrastructure.

### Research Data Management Working Group

UVic's AVP Research Operations, Dr. Rachael Scarth, is currently chairing an RDM Working Group, which will produce an institutional strategy for RDM in response to the forthcoming Tri-Agency RDM policy.

# 2 CHAPTER TWO: RESEARCH METHODOLOGY

## 2.1 Methodological Approach

In order to be better prepared to support RDM practices, the University of Victoria Libraries, with the support of the Office of Research Services, conducted a mixed methods study of researchers from all ranks of faculty members, post-doctoral fellows, and graduate students from across campus. This methodological approach provides a rich source of quantitative and qualitative data that allow for the triangulation of results.

The University of Victoria is part of a number of concurrent survey efforts on post-secondary research campuses across Canada examining the data management practices of researchers. A unique aspect of our study is that librarians participated in both qualitative and quantitative aspects of the study. To do so, they received additional training to acquire the skills necessary to conduct focus group sessions, enabling them to take a leadership role in the data gathering process and work directly with researchers within their subject-liaison areas.

## 2.2 Selection of Participants

Selection of participants for inclusion in the online survey was based on the following criteria: active UVic researchers registered as either faculty members (including, lecturers and instructors, librarians, adjunct professors, assistant professors, associate professors, and full professors), post-doctoral researchers, or graduate students. In-depth interviews were conducted with data-intensive researchers who currently or were previously grant Tri-Agency award holders (Canadian Institute of Health Research (CIHR), the Natural Sciences and Engineering Research Council (NSERC), or the Social Sciences and Humanities Research Council (SSHRC)). This criterion was applied in light of the forthcoming RDM policy from the Tri-Agency that identifies responsibilities of key Canadian stakeholder groups involved in funded research (Government of Canada, 2018). Lastly, a series of focus groups were held with librarians and digital humanities researchers working at UVic Libraries.

## 2.3 Online Survey

The online survey was conducted using the online FluidSurveys platform. The survey instrument was developed by the Canadian RDM Survey Consortium (Portage, 2017) and adopted by the University of Victoria libraries to allow comparative analysis with other institutions across Canada who are apart of the consortium. The survey instrument consists of four main sections: 1) Working with research data; 2) Data sharing; 3) Funder mandates and RDM services; and 4) Demographics and general questions.

## 2.4 Interviews and Focus Groups

Semi-structured personal interviews and focus group sessions with researchers were conducted from June to December 2017. This investigation used an in-depth interviewing snowball sampling approach to data production, so that rather than beginning with a hypothesis, the first step was collecting data through semi-structured face-to-face interviews. Researchers were contacted via invitation letter from their subject-liaison librarians to participate in focus group sessions. Additional focus groups were also conducted with faculty and post-doctoral researchers to collect data on RDM practices and issues. All interviews and focus groups were audio recorded and transcribed, with subsequent text-analysis identifying key points that were marked with a series of codes. Codes were then grouped into similar topics in order to identify major concepts. From these concepts themes emerged, which formed the basis of the theoretical framework related to open data, data curation, and digital scholarship.

## 2.5 Research Ethics and Informed Consent

The informed consent process for this research investigation adhered to the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, Section 2.1 Free and Informed Consent* (Panel on Research Ethics, 2008), as outlined in the *Annotated Guidelines for Completing the Human Research Ethics Board Application for Ethics Approval for Human Participation Research at the University of Victoria* (University of Victoria, 2018). The signed written informed consent form at the end of the study afforded participants the opportunity to exercise their

consent at the conclusion of the study, following debriefing. If a participant expressed concerns about the study, they were given the option of removing their data from the project in the event of perceived or actual conflicts of interest, with the exception of focus group participants whose data was de-identified. Participants were informed of their right to ongoing consent, which included:

- Signing a release/consent form allowing the researcher to use their data at the end of the project;
- Initialling a statement on the consent form signalling their consent to use their data at all stages of the research including transcripts;
- Being informed of their right to withdraw from the research process at any point should an issue arise or to not permit use of their data for certain components of the study; and
- Being informed that names and personal identification would not be associated with the study.

# 3 CHAPTER THREE: QUANTITATIVE SURVEY FINDINGS

The findings of this chapter summarize the responses to the online survey distributed to researchers across the University of Victoria[4]. The survey received a total of 418 responses, approximately 10 % of the total researcher population at UVic[5]. Of the responses, the highest percentage came from graduate students (62.8 %) and professors (32.1 %). Other respondents (15.9 %) included librarians, post-doctoral researchers, sessional instructors and visiting scholars. The following series of figures and tables summarize our findings. Question text is presented (italicized) followed by a breakdown of responses. Note that not every respondent answered every question and therefore the respondent counts for each question vary. As well, where questions allowed for a "check all that apply", response percentages can exceed 100 %.

**Table 1.** Count of survey responses.

| STATUS OF RESPONSES | NUMBER OF RESPONSES | PERCENTAGE (%) |
|---|---|---|
| **COMPLETED** | 349 | 83.5 |
| **PARTIALLY COMPLETED** | 69 | 16.5 |
| **TOTAL** | 418 | 100 |

[4]The survey dataset is available to review online: https://doi.org/10.5683/SP2/1L8NKY
[5]The survey was delivered to 856 Faculty and 3,283 Grad students (n = 4,139)

## 3.1 Characterization of Survey Respondents

This section includes analysis of the responses to survey questions 1-4. Respondents were asked about rank, departmental affiliations as well as funding sources.

Question 1 asked *'Please indicate your rank at UVic'.* Tables 2 and 3 describe the range in research positions of survey respondents. The highest number of responses came from graduate students (62.8 %) and professors (32.1 %).

**Table 2.** Distribution of positions among survey respondents at the University of Victoria.

| | ADJUNCT PROFESSOR | ASSISTANT PROFESSOR | ASSOCIATE PROFESSOR | FULL PROFESSOR | GRADUATE STUDENT | LECTURER/INSTRUCTOR | LIBRARIAN | POST-DOCTORAL FELLOW | OTHER |
|---|---|---|---|---|---|---|---|---|---|
| **% OF RESPONSES** | 0.7 | 5 | 12.7 | 13.7 | 62.8 | 0.2 | 1.2 | 0.2 | 3.4 |
| **# OF RESPONSES** | 3 | 21 | 53 | 57 | 262 | 1 | 5 | 1 | 14 |
| | **Survey Responses** | | | | | | | | |

**Table 3.** Collective positions of survey respondents at the University of Victoria.

| | CONTINUING RESEARCHERS ALL RANKS | GRADUATE STUDENT MA AND PHD | RESEARCH ASSOCIATES POST DOC, LECTURES, VISITING SCHOLARS, LIBRARIANS |
|---|---|---|---|
| **% OF RESPONSES** | 32.1 | 62.8 | 5.0 |
| **# OF RESPONSES** | 135 | 262 | 21 |
| | **Survey Responses** | | |

Question 4 asked *'Which funding sources have you used within the past 5 years? Select all that apply'.* Respondents were asked to select all the funding sources they had used within the past 5 years. They were able to select all that applied including, SSHRC Insight Grant, SSHRC Partnership Grant, SSHRC other (respondent had the opportunity to specify), Canada Council for the Arts, CIHR, CFI, NSERC, ARC (UK), ESRC (UK), EU, Industry, Mellon Foundation, MITACs, NEH (USA), NIH (USA), SSHRC (USA), None, Other (respondent had the opportunity to specify).

**Table 4.** Funding sources for graduate students and professors.

| FUNDING SOURCE | GRADUATE STUDENTS | PROFESSORS |
|---|---|---|
| **SSHRC** | 32 | 52 |
| **CIHR** | 13 | 15 |
| **CFI** | 2 | 21 |
| **NSERC** | 24 | 47 |
| **INDUSTRY** | 3 | 14 |
| **MITACS** | 8 | 20 |
| **OTHER** | 65 | 40 |
| **NONE** | 116 | 11 |

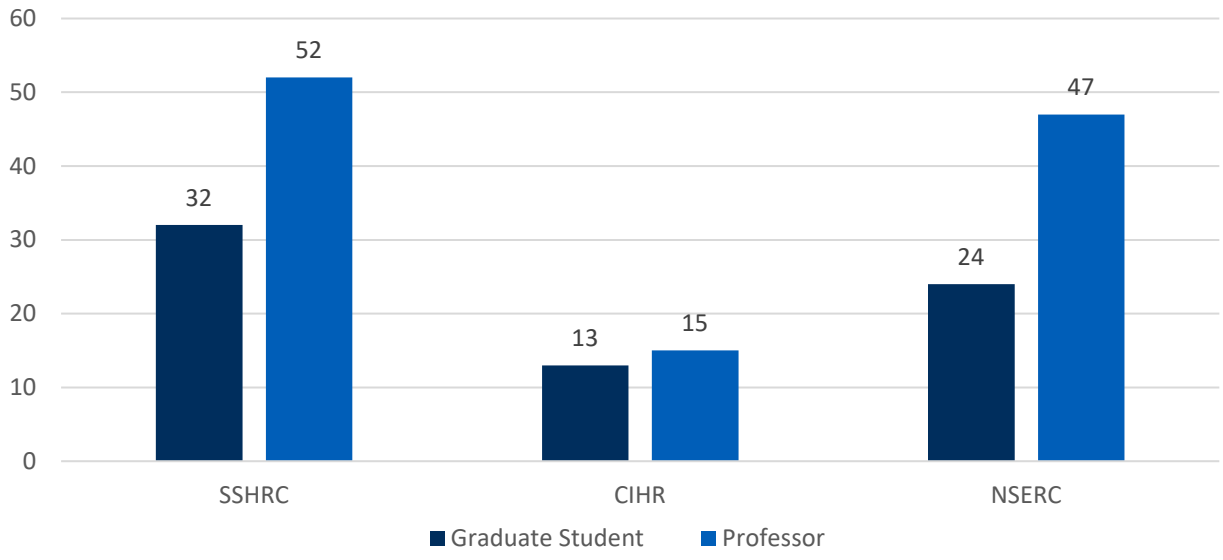



**Figure 2.** Count of Tri-Agency funding recipients, according to major respondent categories.

## 3.2 Summary of Responses

This section includes analysis of the responses to questions 5-15. Respondents were asked questions about their research data, including how they work with them, document them and store them.

### 3.2.1 Types of Research Data

Question 5 asked *'Which of the following best describes the type of research data you generate or use in a typical research project? Select all that apply'.* Data types included "Geospatial" (n= 40), "Instrument specific" (n = 32), "Models" (n= 60), "Multimedia" (n = 126), "Numerical" (n = 148), "Software" (n = 84), "Text" (n = 273), and "Other, please specify" (n = 28). As the respondents were asked to select all that applied, the counts reflect the total number of times each type of data was chosen, for a total of 791 responses.

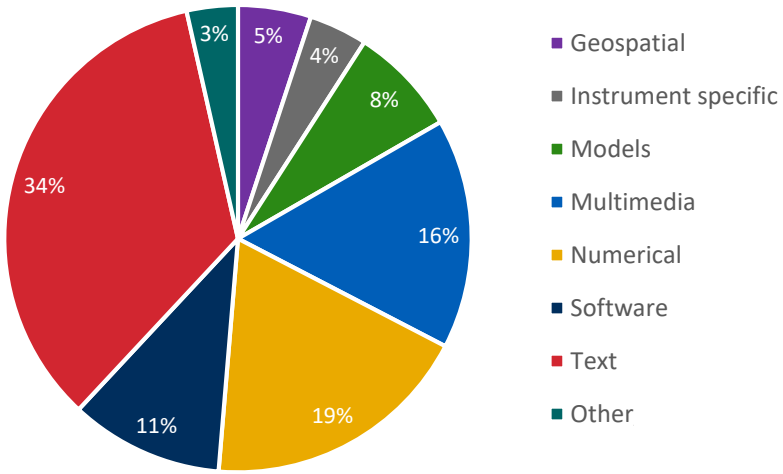



**Figure 3.** Types of data generated or used by respondents in their research projects.

### 3.2.2 Data Storage

Question 7 asked *'Please indicate where you store research data from your current project(s). Select all that apply'.* Counts displayed in Figure 4 represent the total number of times each storage medium was chosen, resulting in a total of 1,342 responses. An important component of good data storage practices involves, not just duplication of data files, but storing copies on multiple storage media. Figures 4 examines the variety of storage options used by respondents, while Figure 5 uses a box and whisker plot to summarize the number of storage media types being used, according to respondent type.

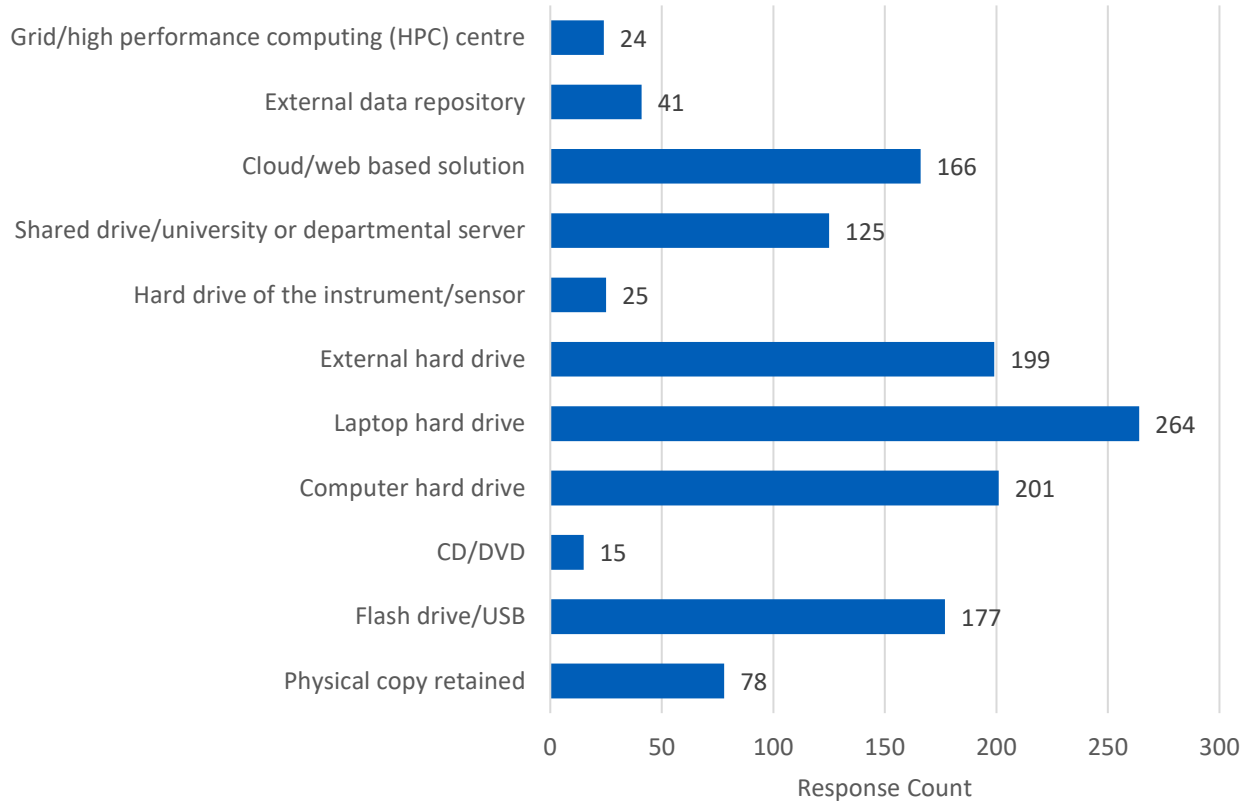



**Figure 4.** Storage media used in current research projects of respondents. Response counts are presented adjacent to each bar.

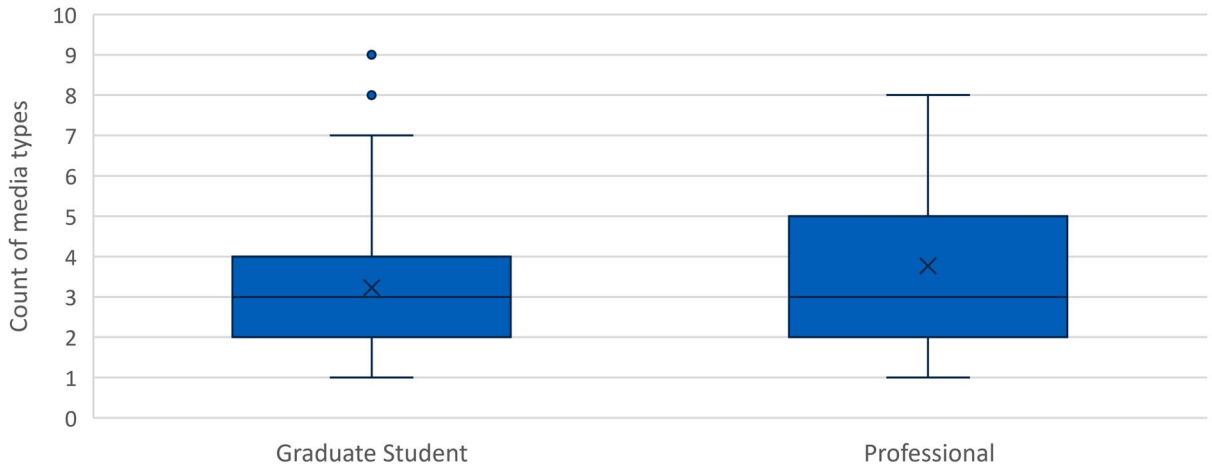



**Figure 5.** Summary of the number of storage media types currently used in research projects, according to respondent type (Graduate Student, n = 243; Professional Researcher, n = 140).

### 3.2.3 Documentation of Data

Question 8 asked *'Do you think there is sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person that is part of your research team to understand and use the research data?'* Respondents reported "Yes" (n = 205), "No" (n = 54) and "Not Sure" (n = 123) to whether another person on the research team could understand and use their data

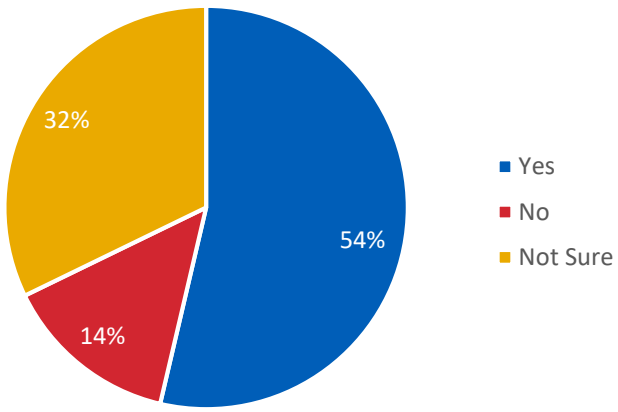



**Figure 6.** Independent understandability of research data for research team members.

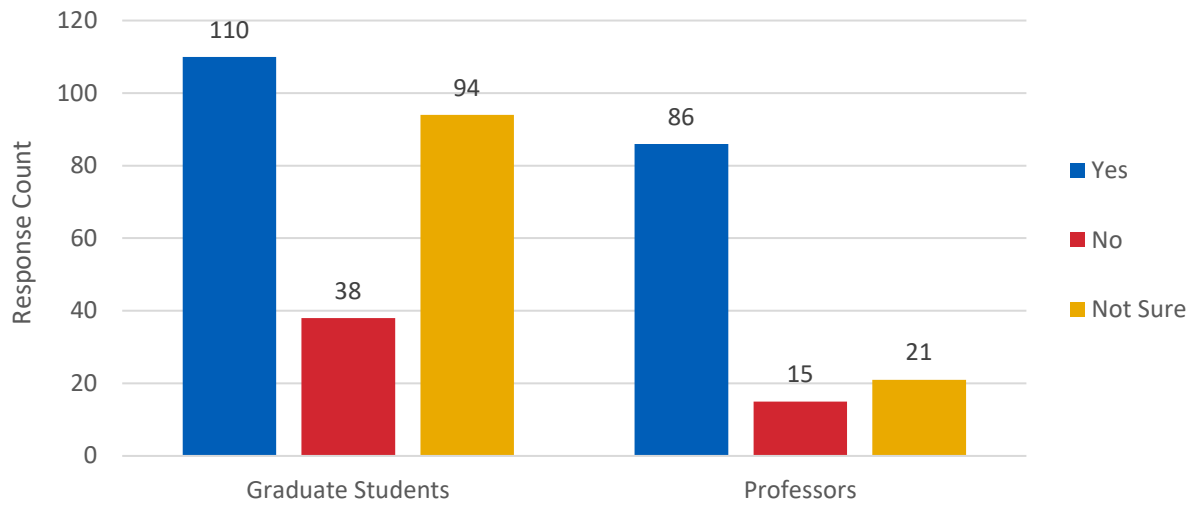



**Figure 7.** Independent understandability of research data for research team members, according to respondent type. Response counts are presented above each bar.

Question 9 asked *'Do you think there is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person that is **NOT** part of your research team to understand and use the research data?'* Respondents reported "Yes" (n = 106), "No" (n = 77), and "Not sure" (n = 199) to whether another person who was not on their research team could understand and use their data.

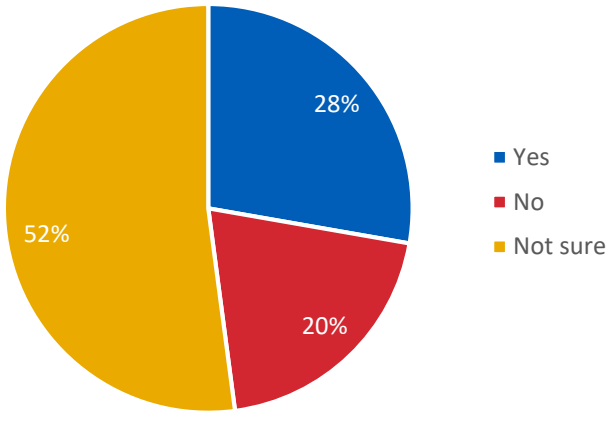



**Figure 8.** Independent understandability of research data for non-team members.

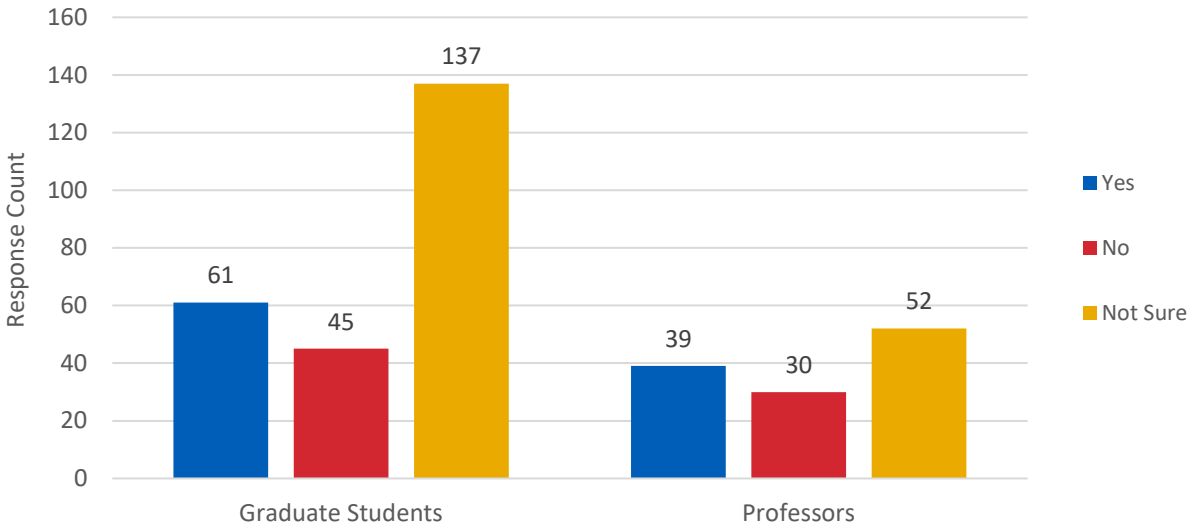



**Figure 9.** Independent understandability of research data for non-team members, according to respondent type. Response counts are presented above each bar.

### 3.2.4 Data Sharing

Question 10 asked *'Which methods of sharing your research data do you currently use? Select all that apply. If you do not currently share your data, choose 'not currently sharing'.*

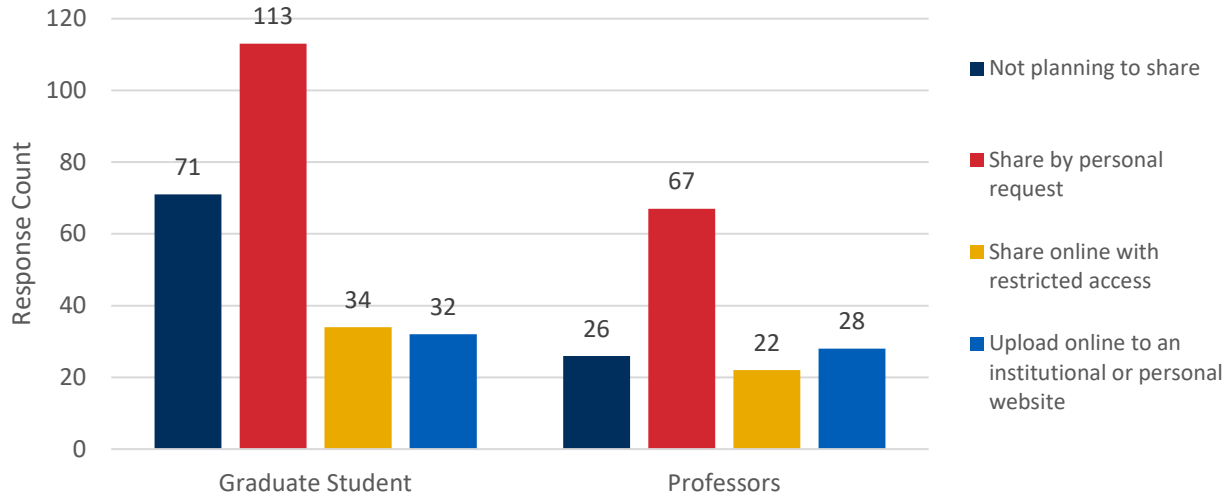



**Figure 10.** Most frequent methods of sharing research data, according to respondent type. Response counts are presented above each bar.

Question 11 asked *'Some research data cannot be shared because of legal or privacy restrictions or embargoes. Which of the following restrictions or embargoes may limit your ability to share your data with others? Select all that apply. If there are no restrictions or embargoes, choose 'there are no restrictions or embargoes on sharing my data with other parties''*.

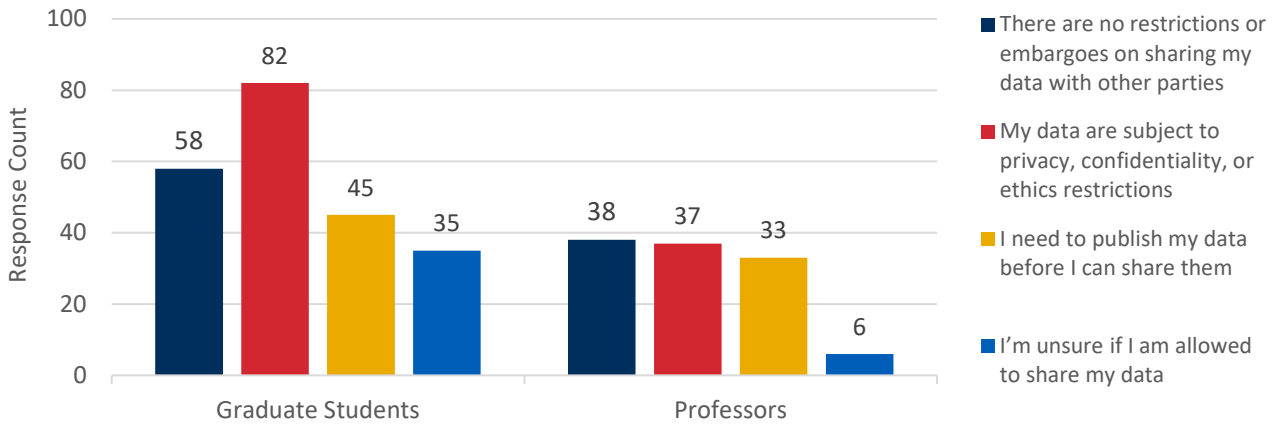



**Figure 11.** Most frequent restrictions on sharing research data, according to respondent type. Response counts are presented above each bar.

Question 12 asked *'What, if any, are the reasons you would not be willing to share your research data and associated methods/tools? Select all that apply. If you are willing to share, choose 'I am willing to share them''*.

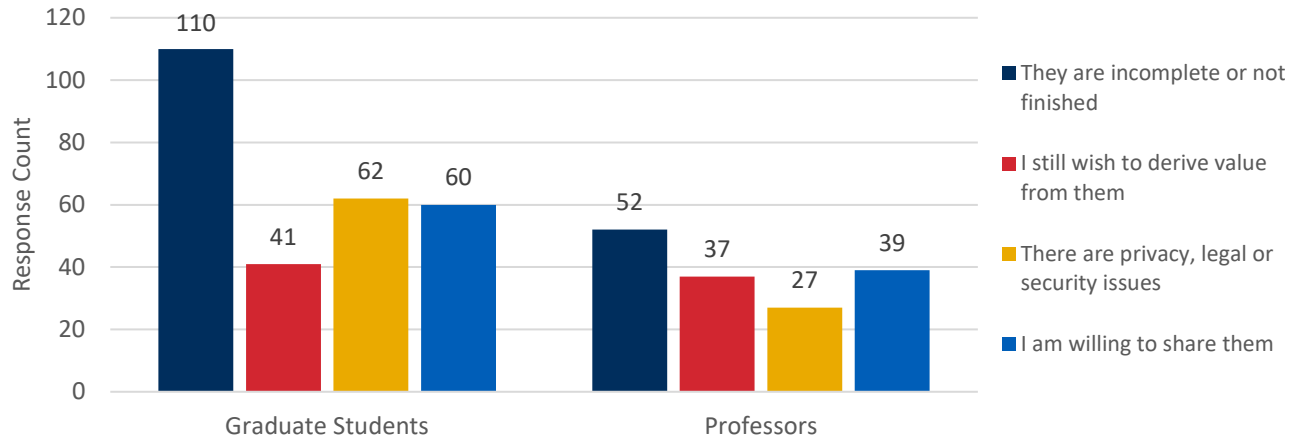



**Figure 12.** Most frequent justifications for not sharing research data, according to respondent type. Response counts are presented above to each bar.

Question 13 asked *'What benefits do you see to sharing your research data? Select all that apply. If you see no benefits, choose 'I see no benefits to sharing my data''*.

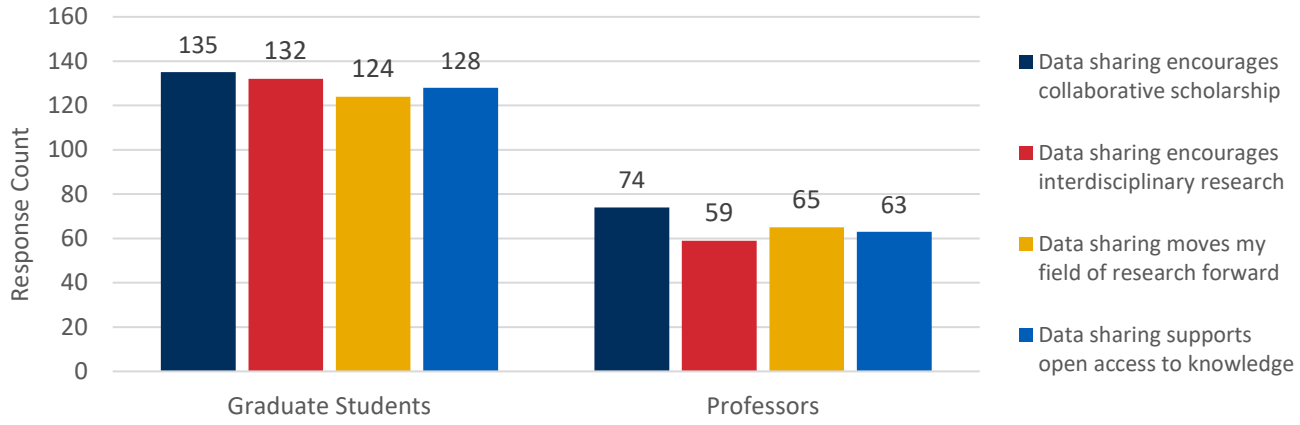



**Figure 13.** Most frequently selected benefits to sharing data, according to respondent type. Response counts are presented above to each bar.

### 3.2.5 Planning and Support

Question 14 asked *'If you were asked to draft a data management plan as part of a grant application, which of the following statements would best describe your situation?'*

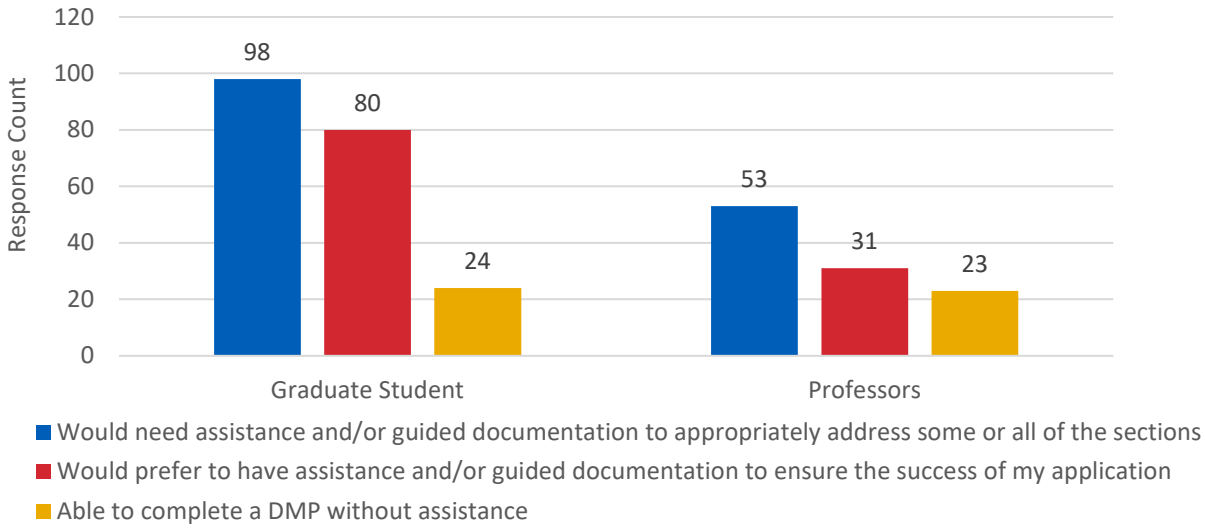



**Figure 14.** Self-assessment of respondents' ability to complete their own data management plans, grouped according to respondent type. Response counts are presented above each bar.

Question 15 asked *'If data management plans were made part of grant applications from funding bodies such as SSHRC, CIHR, and NSERC, how interested would you be in the following services? Please rate your interest in each service'*.

**Table 5.** Interest ratings of respondents in research data management service offerings (see Appendix B for breakdown of responses according to respondent type).

| | RATINGS | | | |
|---|---|---|---|---|
| **SERVICE** | **Very Interested** | **Interested** | **Not Interested** | **Not Applicable** |
| **WORKSHOPS ON BEST PRACTICES IN DATA MANAGEMENT FOR FACULTY** | 70 | 111 | 35 | 106 |
| **WORKSHOPS ON BEST PRACTICES IN DATA MANAGEMENT FOR GRADUATE STUDENTS** | 123 | 142 | 40 | 25 |
| **PERSONALIZED CONSULTATION ON DATA MANAGEMENT PRACTICES FOR SPECIFIC RESEARCH GROUPS OR PROJECTS** | 107 | 136 | 61 | 22 |
| **COMMUNICATION AND INFORMATION ABOUT FUNDING REQUIREMENTS AND JOURNAL REQUIREMENTS REGARDING RESEARCH DATA** | 99 | 162 | 42 | 22 |
| **ASSISTANCE PREPARING DATA MANAGEMENT PLANS TO MEET FUNDING REQUIREMENTS, OR ASSISTANCE CREATING FORMAL OR DOCUMENTED DATA MANAGEMENT POLICIES** | 112 | 153 | 44 | 18 |
| **DIGITIZATION OF PHYSICAL RECORDS** | 73 | 110 | 72 | 62 |
| **ASSISTANCE WITH DOCUMENTING AND DESCRIBING DATA (I.E. METADATA CREATION)** | 77 | 145 | 78 | 21 |
| **ASSISTANCE WITH ISSUES ASSOCIATED WITH DATA PRESERVATION AND/OR SHARING (CONFIDENTIALITY, PRIVACY, ETHICS, LEGAL, INTELLECTUAL PROPERTY RIGHTS)** | 103 | 150 | 57 | 14 |
| **DATA STORAGE AND BACKUP DURING ACTIVE RESEARCH PROJECTS** | 118 | 127 | 64 | 13 |
| **AN INSTITUTIONAL REPOSITORY FOR LONG-TERM ACCESS AND PRESERVATION OF RESEARCH DATA** | 122 | 130 | 57 | 16 |
| **ASSISTANCE WITH DEPOSITING DATA IN APPROPRIATE DISCIPLINARY OR OTHER EXTERNAL DATA REPOSITORIES** | 59 | 150 | 88 | 24 |
| **ASSIGNMENT OF PERMANENT DIGITAL OBJECT IDENTIFIERS (DOIS) FOR DATASETS** | 63 | 135 | 83 | 38 |
| **ASSISTANCE IN FINDING AND ACCESSING DATA SOURCES** | 86 | 145 | 70 | 19 |

## 3.3 Disciplinary Analysis

We explored survey responses according to UVic departmental affiliation, under the assumption that variation exists in how researchers from different disciplinary groups define, collect and curate their research data. For this analysis, the self-declared departmental affiliations of respondents were aggregated into board disciplinary groups using an ontology developed by the Canadian RDM Survey Consortium (see Appendix A). This was done in order to maintain respondent confidentiality, and ensure comparability with the survey data being generated from other Canadian research institutions.

### 3.3.1 Data Storage

The following tables and figures explore the range of storage media used for research data across the disciplinary fields surveyed, pointing to possible differences in requirements of the data generated and distinct practices within disciplinary research cultures.

**Table 6.** Use of media for research data storage by survey respondents, organized by disciplinary affiliation. Percentages of affirmative responses are shown, with total sample sizes (n) presented for each category (ENG = Engineering; SCI = Sciences; ART = Arts and Humanities; BUS = Business; EDU = Education; LAW = Law; SOC = Social Sciences; MED = Medicine and Health Sciences; INT = Interdisciplinary).

| | ENG (n=37) | SCI (n=57) | ART (n=71) | BUS (n=18) | EDU (n=16) | LAW (n=13) | SOC (n=133) | MED (n=29) | INT (n=24) |
|---|---|---|---|---|---|---|---|---|---|
| **FLASH DRIVE/USB** | 35% | 37% | 54% | 33% | 44% | 46% | 45% | 52% | 42% |
| **CD/DVD** | 0% | 2% | 7% | 0% | 6% | 0% | 3% | 3% | 13% |
| **COMPUTER HARD DRIVE (I.E. LOCAL HARD DRIVE)** | 62% | 63% | 41% | 28% | 50% | 38% | 49% | 52% | 42% |
| **LAPTOP HARD DRIVE** | 57% | 72% | 80% | 56% | 75% | 77% | 64% | 66% | 33% |
| **EXTERNAL HARD DRIVE** | 49% | 65% | 62% | 22% | 25% | 69% | 43% | 45% | 38% |
| **INSTRUMENT/ SENSOR HARD DRIVE** | 14% | 14% | 1% | 0% | 6% | 0% | 4% | 14% | 0% |
| **SHARED DRIVE/ UNIVERSITY OR DEPARTMENTAL SERVER** | 30% | 37% | 27% | 17% | 38% | 8% | 29% | 34% | 46% |
| **CLOUD/WEB BASED SOLUTION** | 41% | 37% | 48% | 44% | 44% | 54% | 41% | 41% | 21% |
| **EXTERNAL DATA REPOSITORY** | 22% | 16% | 7% | 6% | 0% | 0% | 7% | 3% | 21% |
| **HIGH PERFORMANCE COMPUTING CENTRE** | 8% | 26% | 0% | 0% | 0% | 0% | 2% | 3% | 8% |
| **PHYSICAL COPY RETAINED** | 5% | 9% | 32% | 0% | 19% | 8% | 23% | 24% | 25% |

Using the data file types respondents reported creating or using as a proxy, we classified respondents as conducting primarily quantitative or qualitative research data. Respondents who reported as not producing any data in text formats were classified as being primarily quantitative research data producers (n=45). Meanwhile, respondents who reported producing data only in text format were classified as being primarily qualitative research data producers (n = 104). Table 7 examines storage media use according to qualitative and quantitative categories.

**Table 7.** Use of media for research data storage by survey respondents, organized according to quantitative and qualitative research categories. Percentages of affirmative responses are shown, with category sample sizes (n) presented.

| | Quantitative Data (n=45) | Qualitative Data (n=104) |
|---|---|---|
| **FLASH DRIVE/USB** | 38% | 46% |
| **CD/DVD** | 0% | 0% |
| **COMPUTER HARD DRIVE (I.E. LOCAL HARD DRIVE)** | 56% | 47% |
| **LAPTOP HARD DRIVE** | 53% | 68% |
| **EXTERNAL HARD DRIVE** | 47% | 43% |
| **INSTRUMENT/ SENSOR HARD DRIVE** | 11% | 0% |
| **SHARED DRIVE/ UNIVERSITY OR DEPARTMENTAL SERVER** | 29% | 17% |
| **CLOUD/WEB BASED SOLUTION** | 31% | 41% |
| **EXTERNAL DATA REPOSITORY** | 13% | 2% |
| **HIGH PERFORMANCE COMPUTING CENTRE** | 13% | 0% |
| **PHYSICAL COPY RETAINED** | 7% | 26% |

The following figures examine the data storage media used by respondents according to disciplinary association, with media classified as being either local storage (Flash Drive/USB, CD, Computer Hard Drive, Laptop Hard Drive, External Hard Drive, Physical Copy) or remote storage (Shared drive/ University or Departmental Server, Cloud/Web Based Solution, External Data Repository, High Performance Computing Centre).

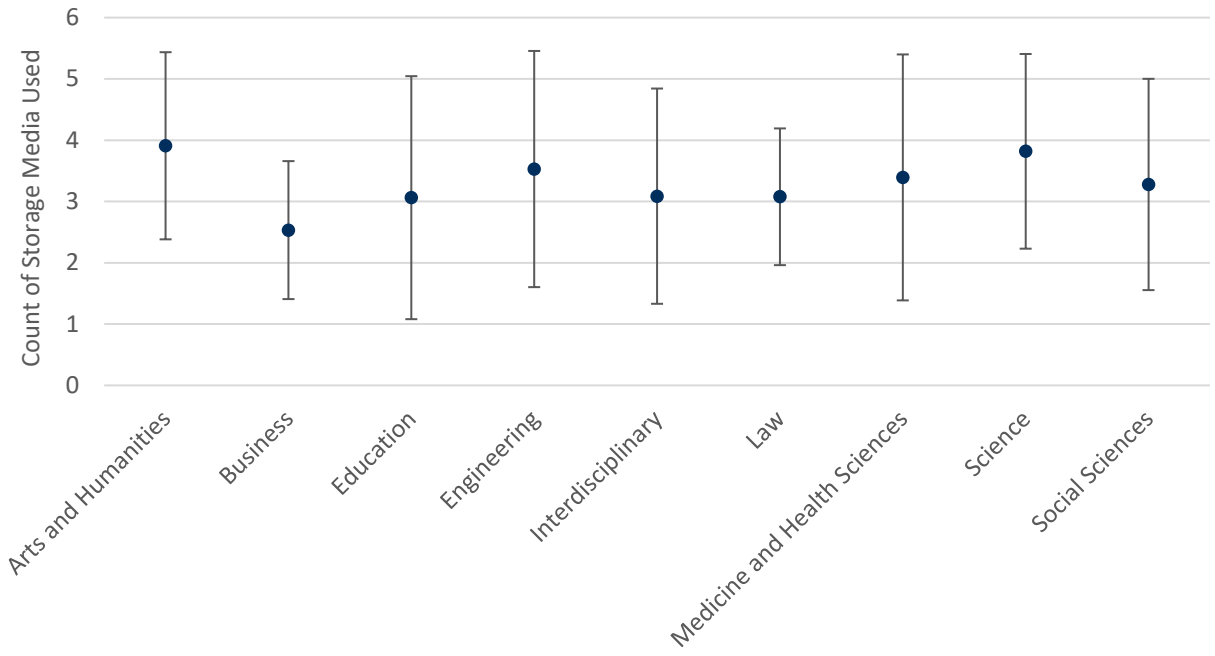



**Figure 15.** Average number of different storage media employed by survey respondents. Error bars measure standard deviation from the mean, indicating the variation in responses within each discipline.

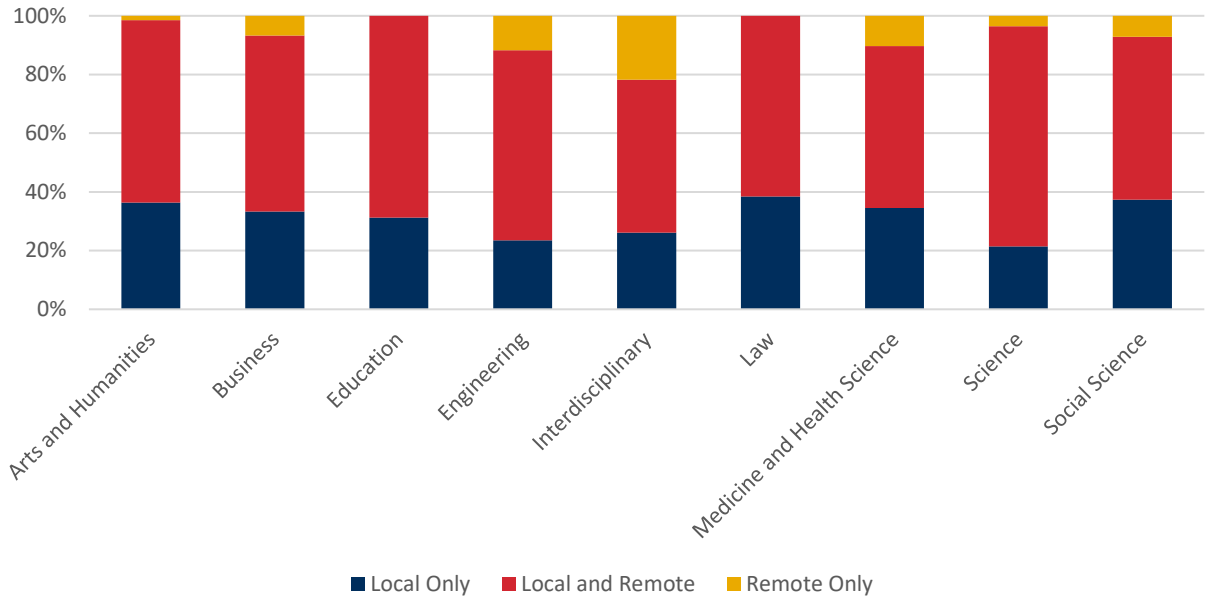



**Figure 16.** Percentage of respondents using local and remote storage options, according to disciplinary affiliation.

### 3.3.2 Sharing Practices

Attitudes towards sharing research data can depend on a range of factors, which may be related to the data itself, including the context in how it was gathered, the processes applied during analysis, and requirements of the intended audience, or which may be the result to established practices and cultural norms that vary across disciplinary areas. The following four tables break out responses to survey questions that deal with attitudes towards data sharing, according to disciplinary groups.

**Table 8.** Data sharing practices of survey respondents, organized according to disciplinary affiliation. Percentages of affirmative responses are shown, with total sample sizes (n) presented for each category (ENG = Engineering; SCI = Sciences; ART = Arts and Humanities; BUS = Business; EDU = Education; LAW = Law; SOC = Social Sciences; MED = Medicine and Health Sciences; INT = Interdisciplinary).

| | ENG (n=37) | SCI (n=57) | ART (n=71) | BUS (n=18) | EDU (n=16) | LAW (n=13) | SOC (n=133) | MED (n=29) | INT (n=24) |
|---|---|---|---|---|---|---|---|---|---|
| **NOT PLANNING TO SHARE** | 24% | 19% | 10% | 28% | 38% | 31% | 29% | 41% | 33% |
| **SHARE BY PERSONAL REQUEST** | 43% | 56% | 56% | 39% | 25% | 31% | 47% | 38% | 42% |
| **SHARE ONLINE WITH RESTRICTED ACCESS** | 11% | 25% | 15% | 17% | 19% | 8% | 13% | 14% | 13% |
| **UPLOAD ONLINE TO AN INSTITUTIONAL OR PERSONAL WEBSITE** | 14% | 14% | 31% | 11% | 25% | 0% | 18% | 3% | 13% |
| **UPLOAD ONLINE TO AN INSTITUTIONAL REPOSITORY** | 0% | 4% | 4% | 0% | 13% | 0% | 5% | 3% | 8% |
| **INCLUDE AS PART OF SUPPLEMENTARY MATERIAL FILES TO A JOURNAL PUBLISHER** | 11% | 26% | 4% | 0% | 0% | 0% | 11% | 3% | 17% |
| **DEPOSIT IN A GENERAL OR DISCIPLINE-SPECIFIC REPOSITORY** | 14% | 35% | 8% | 0% | 0% | 0% | 6% | 7% | 8% |

**Table 9.** Perceived restrictions on sharing research data as reported by survey respondents, organized according to disciplinary affiliation. Percentages of affirmative responses are shown, with total sample sizes (n) presented for each category (ENG = Engineering; SCI = Sciences; ART = Arts and Humanities; BUS = Business; EDU = Education; LAW = Law; SOC = Social Sciences; MED = Medicine and Health Sciences; INT = Interdisciplinary).

| | ENG (n=37) | SCI (n=57) | ART (n=71) | BUS (n=18) | EDU (n=16) | LAW (n=13) | SOC (n=133) | MED (n=29) | INT (n=24) |
|---|---|---|---|---|---|---|---|---|---|
| **THERE ARE NO RESTRICTIONS OR EMBARGOES ON SHARING MY DATA WITH OTHER PARTIES** | 16% | 39% | 42% | 11% | 13% | 15% | 24% | 17% | 4% |
| **I NEED TO PUBLISH MY DATA BEFORE I CAN SHARE THEM** | 38% | 40% | 10% | 17% | 0% | 23% | 15% | 21% | 21% |
| **SHARING MY DATA MAY JEOPARDIZE INTELLECTUAL PROPERTY RIGHTS** | 16% | 16% | 6% | 11% | 6% | 0% | 7% | 7% | 4% |
| **I PLAN TO FILE FOR A PATENT** | 8% | 7% | 0% | 6% | 0% | 0% | 0% | 3% | 0% |
| **MY DATA CANNOT BE SHARED BECAUSE OF COMMERCIAL CONCERNS** | 19% | 4% | 3% | 11% | 0% | 0% | 1% | 7% | 4% |
| **I HAVE A CONTRACTUAL OBLIGATION WITH A THIRD PARTY** | 16% | 16% | 3% | 6% | 0% | 0% | 8% | 3% | 4% |
| **MY DATA ARE SUBJECT TO PRIVACY, CONFIDENTIALITY, OR ETHICS RESTRICTIONS** | 5% | 11% | 15% | 33% | 50% | 31% | 42% | 41% | 58% |
| **MY DATA ARE A MATTER OF PUBLIC SAFETY OR OF A SENSITIVE NATURE** | 0% | 0% | 0% | 6% | 0% | 0% | 6% | 7% | 4% |
| **I'M UNSURE IF I AM ALLOWED TO SHARE MY DATA** | 14% | 4% | 14% | 11% | 13% | 8% | 10% | 10% | 8% |

**Table 10.** Reasons for not sharing research data as reported by survey respondents, organized according to disciplinary affiliation. Percentages of affirmative responses are shown, with total sample sizes (n) presented (ENG = Engineering; SCI = Sciences; ART = Arts and Humanities; BUS = Business; EDU = Education; LAW = Law; SOC = Social Sciences; MED = Medicine and Health Sciences; INT = Interdisciplinary).

| | ENG (n=37) | SCI (n=57) | ART (n=71) | BUS (n=18) | EDU (n=16) | LAW (n=13) | SOC (n=133) | MED (n=29) | INT (n=24) |
|---|---|---|---|---|---|---|---|---|---|
| **THEY ARE INCOMPLETE OR NOT FINISHED** | 30% | 58% | 46% | 11% | 31% | 46% | 39% | 55% | 33% |
| **I STILL WISH TO DERIVE VALUE FROM THEM** | 16% | 28% | 21% | 17% | 6% | 23% | 17% | 24% | 17% |
| **I DO NOT HAVE THE TECHNICAL SKILLS OR KNOWLEDGE** | 3% | 2% | 10% | 0% | 0% | 0% | 5% | 17% | 0% |
| **I DO NOT HOLD THE RIGHTS TO SHARE THEM** | 14% | 19% | 14% | 0% | 13% | 8% | 18% | 10% | 8% |
| **FUNDING BODY DOES NOT REQUIRE SHARING** | 8% | 5% | 3% | 0% | 0% | 15% | 2% | 3% | 4% |
| **I BELIEVE THEY SHOULD NOT BE SHARED** | 5% | 2% | 7% | 11% | 13% | 23% | 8% | 0% | 8% |
| **I DID NOT KNOW I COULD SHARE THEM** | 3% | 7% | 6% | 0% | 0% | 15% | 2% | 0% | 4% |
| **INSUFFICIENT TIME** | 22% | 21% | 25% | 11% | 19% | 8% | 9% | 3% | 17% |
| **LACK OF STANDARDS TO MAKE THEM USABLE BY OTHERS** | 16% | 16% | 6% | 17% | 6% | 0% | 6% | 10% | 8% |
| **LACK OF FUNDING** | 19% | 11% | 17% | 6% | 6% | 8% | 8% | 14% | 25% |
| **NO PLACE TO PUT THEM** | 3% | 11% | 4% | 0% | 13% | 8% | 6% | 7% | 8% |
| **THEY ARE NOT USEFUL TO OTHERS** | 5% | 5% | 11% | 11% | 6% | 8% | 6% | 7% | 0% |
| **THERE ARE PRIVACY, LEGAL OR SECURITY ISSUES** | 14% | 7% | 7% | 33% | 50% | 31% | 27% | 38% | 50% |
| **MY DATA COULD POTENTIALLY BE USED WITHOUT PROPER CITATION** | 8% | 18% | 6% | 17% | 6% | 15% | 3% | 7% | 21% |
| **I'M CONCERNED MY DATA COULD BE USED WITHOUT PROPER CITATION OR ACKNOWLEDGEMENT** | 16% | 30% | 17% | 17% | 19% | 15% | 6% | 21% | 21% |
| **I AM WILLING TO SHARE THEM** | 24% | 35% | 32% | 6% | 25% | 8% | 28% | 17% | 8% |

**Table 11.** Perceived benefits to sharing research data as reported by survey respondents, organized according to disciplinary affiliation. Percentages of affirmative responses are shown, with total sample sizes (n) presented (ENG = Engineering; SCI = Sciences; ART = Arts and Humanities; BUS = Business; EDU = Education; LAW = Law; SOC = Social Sciences; MED = Medicine and Health Sciences; INT = Interdisciplinary).

| | ENG (n=37) | SCI (n=57) | ART (n=71) | BUS (n=18) | EDU (n=16) | LAW (n=13) | SOC (n=133) | MED (n=29) | INT (n=24) |
|---|---|---|---|---|---|---|---|---|---|
| **I SEE NO BENEFITS TO SHARING MY DATA** | 14% | 4% | 4% | 0% | 6% | 8% | 8% | 7% | 8% |
| **DATA AVAILABILITY PROVIDES SAFEGUARDS AGAINST MISCONDUCT, DATA FABRICATION AND FALSIFICATION** | 11% | 39% | 27% | 17% | 19% | 15% | 26% | 48% | 17% |
| **DATA SHARING AND/OR REPLICATION STUDIES HELP IN THE TRAINING OF NEXT GENERATION RESEARCHERS** | 27% | 47% | 39% | 22% | 44% | 23% | 49% | 48% | 29% |
| **DATA SHARING ENABLES MY DATA TO BE CITED AND INCREASES MY RESEARCH IMPACT** | 24% | 49% | 48% | 28% | 44% | 31% | 34% | 31% | 46% |
| **DATA SHARING ENCOURAGES COLLABORATIVE SCHOLARSHIP** | 38% | 61% | 56% | 33% | 56% | 54% | 59% | 52% | 54% |
| **DATA SHARING ENCOURAGES INTERDISCIPLINARY RESEARCH** | 30% | 53% | 55% | 33% | 63% | 54% | 53% | 45% | 58% |
| **DATA SHARING MOVES MY FIELD OF RESEARCH FORWARD** | 30% | 60% | 56% | 22% | 56% | 46% | 48% | 52% | 58% |
| **DATA SHARING REDUCES REDUNDANT DATA COLLECTION** | 19% | 51% | 34% | 33% | 38% | 23% | 39% | 38% | 25% |
| **DATA SHARING SUPPORTS OPEN ACCESS TO KNOWLEDGE** | 35% | 63% | 56% | 22% | 56% | 31% | 51% | 55% | 42% |
| **RE-ANALYSIS OF DATA HELPS VERIFY RESULTS** | 24% | 53% | 31% | 33% | 31% | 31% | 38% | 52% | 29% |
| **WELL-MAINTAINED DATA HELPS RETAIN DATA INTEGRITY** | 8% | 47% | 28% | 22% | 31% | 15% | 34% | 38% | 29% |

### 3.3.3 Data Management Planning

Researchers from certain disciplines may already be familiar or knowledgeable about data management planning, because of expectations set by domain-based research associations and granting agencies. In order to predict demand for data management planning support, we were interested in determining whether differences in perceived abilities to complete a data management plan existed between disciplinary groups surveyed. Table 12 summarizes survey respondents' self-ratings of their ability to complete data management plans for their research, with or without additional support, according to disciplinary affiliation.

**Table 12.** Self-assessment of respondents' ability to complete their own data management plans, grouped according to disciplinary affiliations. Percentages of affirmative responses are shown, with total sample sizes (n) presented.

| | REQUIRE SUPPORT TO COMPLETE A DMP | ABLE TO COMPLETE A DMP, BUT WOULD PREFER SUPPORT | ABLE TO COMPLETE A DMP WITHOUT ASSISTANCE |
|---|---|---|---|
| **ENGINEERING (n = 37)** | 35% | 22% | 16% |
| **SCIENCE (n = 57)** | 28% | 42% | 21% |
| **ARTS & HUMANITIES (n = 71)** | 45% | 25% | 4% |
| **BUSINESS (n = 18)** | 39% | 11% | 17% |
| **EDUCATION (n = 16)** | 63% | 19% | 0% |
| **LAW (n = 13)** | 38% | 23% | 8% |
| **SOCIAL SCIENCES (n = 133)** | 36% | 31% | 17% |
| **MEDICINE & HEALTH SCIENCE (n = 29)** | 48% | 34% | 3% |
| **INTERDISCIPLINARY (n = 24)** | 33% | 29% | 13% |

## 3.4 Survey Highlights

**Data Management Planning**

Additional support for completing data management plans is either desirable or deemed necessary according to the vast majority of respondents across all disciplines. Generally, this finding was consistent across most disciplines. Respondents from sciences were most confident in their ability to create a data management plan, whereas respondents from education-related disciplines were most likely to indicate greater need for additional support.

**Data Documentation**

The majority of researchers surveyed, indicated that they do require guidance or assistance in documenting and describing their data. This was shown by the majority of respondents who did not believe, or were unsure if there is sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person outside their lab to understand and use their research data. This finding was also consistent across disciplinary groups.

**Data Storage and Security**

Respondents across all disciplines are employing multiple media to store their data, and most use a combination of local and remote storage (note that this doesn't necessarily mean all of their data is backed up to multiple media, but it is at least indicative that this could be possible).

**Data Sharing**

Most respondents agree there are benefits to sharing research data. Respondents most frequently agreed with data sharing benefits related to collaborative scholarship and interdisciplinary research. Faculty respondents were also likely to see sharing as leading to supporting open access to knowledge, while graduate students also saw sharing as important for driving research progress. In the breakdown by disciplines, a few differences emerged. For

example, respondents from both medicine and science disciplines were more likely to view sharing data as a safeguard against misconduct, fabrication and falsification, and important for verification of results through re-analysis. Meanwhile, respondents from business and engineering were less likely to agree with the listed statements of data sharing benefits.

In terms of existing restrictions to sharing, some disciplinary differences emerged in the responses. Respondents from arts and humanities, and the sciences were more likely to indicate there were no restrictions or embargoes in sharing data with other parties. Meanwhile, respondents from engineering and sciences were more likely to report needing to publish data before being able to share them. Respondents from education, social sciences, medicine, business and law more often reported their data being subject to privacy, confidentiality and ethics restrictions as a barrier to sharing, who also frequently cited privacy, legal or security issues as reasons for not sharing. Commercial concerns emerged as unique barriers to sharing data for business and engineering, compared to other disciplinary groups.

The most frequently selected reason for not sharing data was incomplete data (a possible artifact of the large proportion of graduate student respondents). Insufficient time, wanting to continue deriving value from research data, and concerns with improper citation and acknowledgement also emerged as frequent responses across disciplinary groups.

**Research Data Services**

Respondents showed interest in all research data services queried, with responses highest for:

1. Workshops on best practices in data management for graduate students
2. An institutional repository for long-term access and preservation of research data
3. Data storage and backup during active research projects
4. Assistance preparing data management plans to meet funding requirements, or assistance creating formal or documented data management policies

# 4 CHAPTER FOUR: QUALITATIVE RESEARCH FINDINGS

Twenty-three in-depth interview sessions were held with graduate students and faculty researchers from across the University of Victoria between June 2018 to September 2018. Three focus group sessions were also held with researchers and librarians at the University of Victoria from June 2017 to August 2018. Twenty-one participants attended these focus groups sessions. Interview and focus group participants were selected from the following departments: business, digital humanities, education, fine arts, human and social development, humanities, law, libraries, science, and social sciences. The following chapter highlights some of the key findings of these conversations.

**Table 13.** Count of interview and focus group participants.

| STATUS OF RESPONSES | NUMBER OF PARTICIPANTS | PERCENTAGE, % |
|---|---|---|
| **IN-DEPTH INTERVIEWS** | 23 | 52.2 |
| **3 FOCUS GROUPS** | 21 | 47.7 |
| **TOTAL** | 44 | 100 |

**Table 14.** Focus group participants.

| DATE | FOCUS GROUP | NUMBER OF PARTICIPANTS |
|---|---|---|
| **JULY 26, 2017** | Librarians | 7 |
| **AUGUST 24, 2017** | Librarians | 5 |
| **SEPTEMBER 25, 2017** | Digital Humanities Researchers | 9 |
| **TOTAL** | | 21 |

This research process was participatory in nature. In-depth qualitative interviews and focus group sessions ranged from 20 to 90 minutes. Sessions were audio recorded, transcribed, and analyzed using NVivo version 11.0 qualitative coding software.

## 4.1 Core Themes

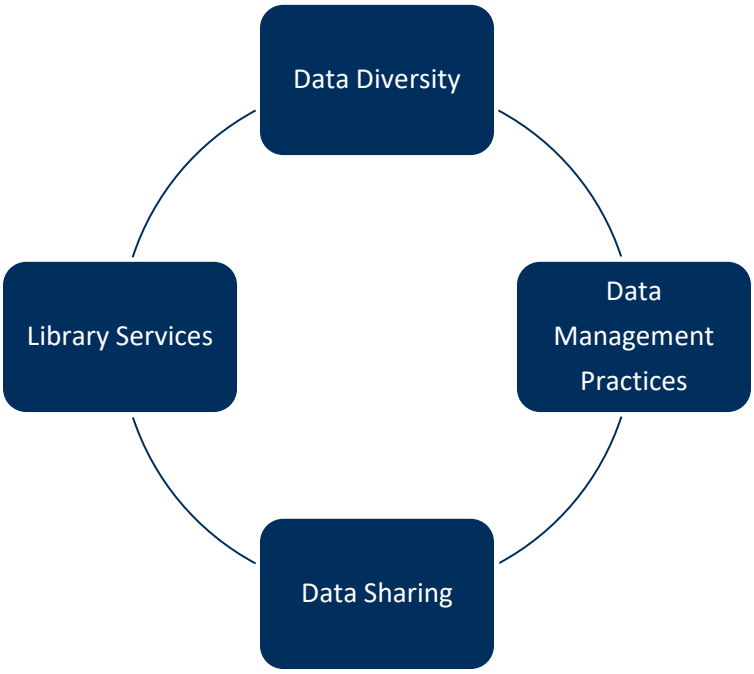



**Figure 17.** Core themes identified through interviews and focus groups.

Conversations centered around four themes: 1) Data Diversity; 2) Research Data Management Practices; 3) Data Sharing, and 4) the Role of Library Support Services.

By engaging the research community through this process, we were able to surface questions and concerns about the draft Tri-Agency RDM policy, and identify barriers and challenges to data management and open data sharing. The following sub-sections in this chapter outline in more detail important aspects of each of the four core themes of the study.

## 4.2 Data Diversity

Research data can be simply defined as the original sources or material that have been created for a given research project in digital or non-digital formats. In the Tri-Agency's Statement of Principles on Digital Data Management, research data are described as follows:

> *"Research data include observations about the world that are used as primary sources to support scientific and technical inquiry, scholarship, and research-creation, and as evidence in the research process. Research data are gathered through a variety of methods, including experimentation, analysis, sampling, and repurposing of existing data. They are increasingly produced or translated into digital formats. When properly managed and responsibly shared, these digital resources enable researchers to ask new questions, pursue novel research programs, test alternative hypotheses, deploy innovative methodologies, and collaborate across geographic and disciplinary boundaries."* (Government of Canada, 2016)

In practice, defining research data is not so straightforward, and there is considerable variation in vocabulary across disciplines at the University of Victoria. Research data have different meanings depending on numerous factors, which may include the faculty or discipline, the research methods applied, the subject matter, and its intended uses. Figure 18 lists the some of the ways in which interview and focus group participants described research data.


**Figure 18.** Defining research data.

One researcher from the humanities explained "*research is collecting information and processing information in order to create written documents*" while another added that it is a way to "*share knowledge with my colleagues so be able to produce presentations and workshops. Research is being transferred into more of a community practice*." This notion of a community of practice is in line with how one Indigenous researcher explained the concept of research data as "*ways that our community, kind of relates to the world, so it's our worldview, it's got to be embedded in the sense of language, ceremony, the land or the water that you're in, and our histories so those interrelationships are key to I guess what you think of especially in Indigenous forms of knowledge or data*."

These descriptions are very different in comparison to the ways in which some quantitative researchers described research data. One economist explained that research data is "*economic measures, with a lot of co-variants like cultural co-variants [is the] data we collect*." While a mathematician stated, "*in pure mathematics that sort of data doesn't really exist. It's formulae and theorems.*"

**Table 15.** Concepts of research data, organized by faculty.

| FACULTY | EXPLANATION |
|---|---|
| **SOCIAL SCIENCES** | "a presentation or a hand out"<br>"data tends to flow around narrative, social narrative"<br>"data is kind of that window into what we need for evidence based decision making globally"<br>"a lot of graphs and images, a lot of plots, well lots of tables, parameters over time" |
| **HUMAN AND SOCIAL DEVELOPMENT** | "creating things to then give back to the community"<br>"window into how we can look at what is going on in the world"<br>"data that comes out of those stories, so a lot of it's going to be qualitative and some quantitative, possible application and then I think the other forms of data that I think of is data might be maps might be songs" |
| **HUMANITIES** | "a manila folder in a filing cabinet and that's our data".<br>"we're creating image files, we're creating PDF's, we're creating XML, we're creating websites, etcetera, but it's still flat data"<br>"land title searches, thousands of transfer of title documents that we've used to chronical individual histories or properties, and so those take a very different data form"<br>"you have primary sources, your primary data, you have secondary data which is intersectional of primary sources, whatever that looks like and then you would have tertiary data, which are the finding aids, totals, maybe Metadata catalogs, that sort of thing that can be derived from those materials" |
| **ENGINEERING** | "they're all digital, they're numbers, organize it all into images but the data is all numeric"<br>"raw data which may be coming from so again my example of buildings or structures" |
| **EDUCATION** | "the data is more kind of qualitative, seeing, observing, participating"<br>"they would be images, manuscripts in a printed book, descriptions of those images" |
| **SCIENCE** | "It's unprocessed data, processed data, images, continuous live data coming through on time, body composition over time" |

| | |
|---|---|
| | "numerical data, most certainly we work with the numerical data to produce tabulated data and graphs"<br><br>"spreadsheet that says the field site, what the characteristics of the field site, what date it was collected for how long and the data is there you can download it, it will give you and idea about contaminants" |
| **FINE ARTS** | "what you observe and what you participate, and what you read and read and counter read"<br><br>"it's everything, knowledge, information, stories, gossips, rumours, songs"<br><br>"there's photographs, there's videos, there's art work. There's many different kinds of audio recordings, there's playbills from theatre performances, covers, t-shirts" |
| **BUSINESS** | "any cached data that I pull off online discussions or screen shots, or copies of threads, or that sort of thing all that counts as data"<br><br>"email correspondences" |

The diversity of interpretations and descriptions of research data becomes apparent through reading the comments listed in Table 15. When asked what research data are, the researchers in our study explained that research data varies depending on the discipline in which it is generated and used, and there was still considerable variation in defining research data between and within the faculty groups surveyed. This variation is related not only to discipline, but also the methodological framework being applied, for example, qualitative, quantitative, mixed methods, Indigenous methodologies, etc. Therefore, prior to setting RDM requirements or expectations, whether in line with the Tri-Agency or otherwise, there should be a clear and consistent understanding of digital research data, as a way evaluating whether expectations and requirements are being achieved.

## 4.3 Research Data Management Practices

A solid understanding of researcher practices throughout all the stages of a research project is important for establishing data management practices that improve data accessibility and reusability over time. Thorough our conversations with researchers, we gained a deeper understanding and appreciation for the diversity of the ways in which researchers manage, curate and share their data. These practices vary according to not only a researcher's discipline, but also their stage of professional development (e.g. faculty researcher or a graduate student).

Many researchers rely on various software to support them in the process of organizing and manipulating their research data. For qualitative data, some researchers indicated using programs such as NVivo and ATLAS.ti to manage and curate their data. One social sciences researcher commented, "*We store most of our data, especially the text, in version control software.*" Meanwhile, researchers generating quantitative data identified programs such as R, SPSS, Excel, or Matlab when describing how they manage and manipulate their data.

Regarding storage, many researchers described storing data on external platforms like Google Drive. One research in the humanities explained that they use eternal resources to curate and manage their research data because "*I don't have space to curate it anywhere… if I can't store everything with me at the same time I have to go to third party suppliers because my UVic account [is too small]*". Another researcher in the social sciences expressed a similar perspective "*we only have a gig of backup space available to us. I don't feel like I have access to sufficient storage*". Storage space is an important issue in data management. A researcher in the faculty of human and social development further expressed concerns and fears of losing their data based on their inability to store large amounts of data. "*First I went with Dropbox, but then stuff started getting lost so I didn't trust it any more so I got Google Drive and again very recently all my Google Drive stuff disappeared and it was all in my trash because of a software glitch between the software on my computer and the software on cloud it just decided to throw everything away…I have personally sought out third party suppliers to allow me to live, if I had money I would buy a computer with a large enough hard drive.*"

While some researchers rely on third party software to support them with data storage and curation, others have devised their own research systems to help them keep track of their data. These researchers will use their own computers and departmental servers to store their data, but have devised their own strategies for data curation. One social science researcher stated that "[I] *insist the students save everything to a CD-ROM"*, while another mentioned that *"I came up with this 1, 2, 3, system where 1 is stuff other people give us, 2 is anything that we do and 3 is things that we share.*" Another social science researcher explained, "*Within every folder*

*there's data files describing all the architecture standardized codes, so it's a little bit like a Dewey Decimal System for our data. Data archives is the core thing I manage and this is where all my students have their projects basically and within data archive, this is how I describe things special data, model data, field data, completed projects.*" A researcher from fine arts described that "[I'm] *always thinking about it and always looking for systems that make most ready those virtues of data management that is not lost, you don't need to have a very complex system but you do need to have a system that will collect things, that will gather things, that wont lose things, that wont lose things out of carelessness or maybe accidental overriding but also wont lose things if your house burns down.*" While many researchers indicated a preference for managing their own data, many also recognized that some their data could be stored by the UVic Libraries. One education researcher mentioned that "*we handed over that to the library because they have an AToM post.*"

## 4.4 Data Sharing

Understanding how researchers share their data is critical to ensuring appropriate upstream data management practices are applied. Through our conversations, we gained insight into the data-sharing culture among researchers at the University of Victoria. Our findings show that many researchers are open to sharing their data by personal request with their research community, with the understanding that this will advance knowledge in their disciplines. However, many challenges and barriers to data sharing were also identified in our conversations.

In our conversations, researchers regarded the cost of data curation as an important barrier to sharing, particularly in relation to the financial burden to utilize and maintain access to necessary software, and the storage cost needed to securely store large data files for long duration.

Researchers expressed caution about the risk of gambling with the long-term survival of software, with some raising questions, such as "*will the software become obsolete over time?*" and "*what will happen to my data over time?*" Researchers also discussed their transition into

retirement and were cognisant of the need to have a dependable repository that will preserve their data while providing access back to community, the public, and other researchers.

Dependence on commercial software for data analysis and management in many research disciplines was discussed as a barrier for sharing and long-term access. Relying on proprietary software also limits how widely resulting data can be shared, given that all potential users would also need access to the same technology. Over time, a researcher's ability to access to their own data may be limited by reliance on proprietary software, particularly if access is licensed through an institutional subscription, over which they have less control. As well, over time, software and related file formats become obsolete. If researchers do not take active measures to convert their data to more open and accessible formats, they risk their data becoming unusable. The use of commercial software was also viewed as potentially problematic for some research communities where projects are bound to a sense of "relational accountability', which is an ethical responsibility to uphold community expectations. For example, when working with Indigenous communities, one researcher explained, "*there are multiple ethical standards that are being navigated and data curation in a commercial software with proprietary rights may not be conducive to that agreement.*"

Participants discussed their concerns with data privacy, particularly for sensitive data that are subject to additional ethical, legal, and commercial considerations. How and when sensitive data can be shared is often limited by ethical agreements made with communities and individuals, or even legal agreements made, for example, with commercial entities. Some of the researchers interviewed work with vulnerable groups and acquire personal data that must remain confidential until it has been anonymized and de-identified. In these types of situations, researchers discussed their concerns with the length of time required to curate data, while ensuring privacy and security standards. So even while researchers strive to share their data, there are important ethical considerations that constrain open access and data sharing. Similarly, some researchers also discussed a reluctance to curate data in open systems or in software that

was not backed-up to a personal server because of identification concerns and breaches of confidentiality.

Beyond barriers set by cost, technology, and data confidentiality, existing cultural norms may limit data sharing. Some researchers we spoke to expressed concerns that shared data may be misinterpreted and used inappropriately to produce incorrect narratives and false conclusions, highlighting anxiety about working in a more open environment. Some participants also expressed that they were not comfortable sharing data until its ability to be used for their own publications and other scholarly outputs has been exhausted. This may result from the "publish or perish" culture that is a reality in certain disciplines, and highlights challenges with the academic reward system, which should be addressed to value well-curated research data as an important scholarly output and encourage data sharing.

## 4.5 The Role of Library Services

The University of Victoria Libraries' current strategic directions identify three core principles:

1. **Open:** UVic Libraries will connect people, knowledge, and expertise through partnerships and collaborations, as well as create open avenues to research and to physical and virtual spaces.

2. **Engaged:** UVic Libraries will be an active collaborator and connector to enhance the learning, teaching, and research activities of the University, and embrace its role as an access point to the University for the broader community

3. **Enduring:** UVic Libraries will focus on developing long term, flexible, nimble, and durable approaches to its role as a facilitator of student and faculty success. The Libraries will enhance the vibrancy of the local, regional, and global communities with which it engages.

The findings from this study shed light on how the library can apply these principles to supporting researchers in the digital environment at the University of Victoria. There is a clear demand among researchers for greater understanding of RDM and related training and services. Researchers expressed interest in receiving assistance in a range of areas from data management planning, documentation and description, storage and access, and preservation. Identifying key service areas is an important step in positioning UVic Libraries to provide support to researchers going forwards. Participants discussed the library's role in supporting RDM according for five key areas, which provide insight into how the researchers in our study view the important role that the library can play in supporting RDM and encouraging data sharing in the community.

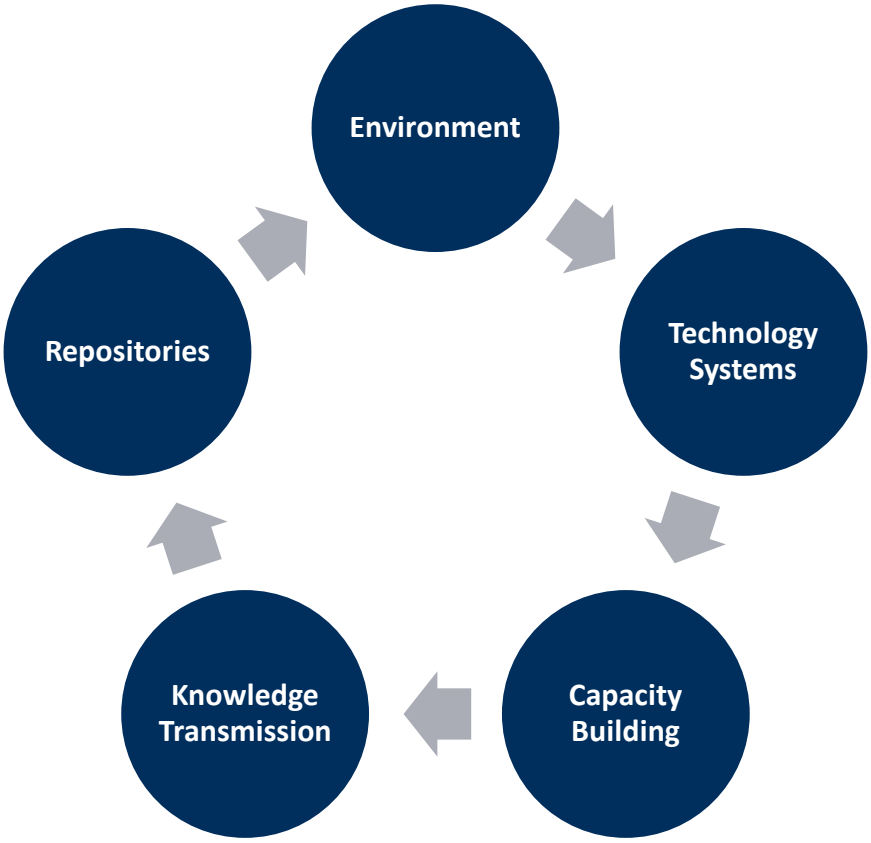



**Figure 19.** The role of library services in supporting research data management.

**Environment**

In our discussions, the library was characterized as a place that fosters a sense of community among researchers, and supports interdisciplinary collaboration between researchers and the university. As one education researcher suggested, "*That's a lot of data, so hence my approach to the library to see whether these data can be made available for other people to use. It seems to me, it would be a perfect role for the library to maintain it*". The library can play a role in creating an environment and culture that adapts and responds to the needs of researchers, offering advice, knowledge mobilization, and technological systems that support data storage and access over the long term.

**Technology Systems**

Technology access is a significant service of the library to aid researchers in the management of research data. Library systems can offer support at various points of the research lifecycle. The digitization of data was identified as a key step in digital scholarship for library services. The curation steps applied to data before depositing to a repository, including the conversion of file formats and application of metadata standards, were viewed as key actions the library could play to improve the accessibility and discoverability researcher data. Participants discussed how access and technological support for software such as Zotero, AtOM, and Jstor were paramount to their successful data management. This regime could be complimented with additional software support in the library for more discipline-specific tools. For instance, several participants discussed how a GIS specialist was helpful for students and a way to support further engagement in using library services. Participants also explained how an important part of their data management plan was to pass over their datasets to the library upon completion, and some senior researchers expressed a desire to transfer data files to the library as they begin to approach retirement. Overall, the technical services offered through the library were considered important for the long-term accessibility and usability of research data. As well, the library's services and tools were regarded as valuable alternatives to proprietary software and commercial services that pose some challenges, as discussed in the preceding section.

**Building Capacity**

Participants discussed their visions for library support services in terms of a liaison model with researchers. The sponsorship of research through partnerships with librarians and other groups on campus was identified as a way to start building research capacity across campus. The role of libraries in supporting connections to other research groups was also discussed.

**Knowledge Transmission**

Participants viewed education for the academic community about library services as an important service itself. Through short courses and workshops, the library could support students and researchers on an on-going basis to develop strong data management plans, curate their data, and publish data in secure repositories. The need for a range of services offered by the Library varies across and within each faculty. These differences discussed elsewhere in this report are important and could provide direction for related outreach.

**Repositories**

Researchers expressed concerns about their data becoming at risk to loss, theft, hardware failure, and other threats over time, highlighting demand for repository platforms where data can be stored and maintained. Participants expressed interest in the library facilitating access to reputable data repositories that provide secure access and long-term preservation of research data. One social sciences researcher expressed interest in the library publishing electronic data sets as a cost effective and efficient service for researchers. Another researcher suggested "*a repository could be developed at the national level, reducing the duplication of raw data and promoting a larger community of research*." A reliable search mechanism for repositories was also suggested, allowing frequency of access to be measured and research impact thereby measured. The library does currently offer researchers access to such to a multidisciplinary data repository platform, and the development of a national research data repository is currently being developed with the support of the national research library community. UVic Libraries should continue to develop their capacity to support researchers with these and other repository platforms, including domain-specific repositories.

# 5 CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

To gain a better understanding of current RDM practices and anticipate future needs, this study employed a mixed methods approach to investigate how researchers at the University of Victoria manage their research data. Data were acquired through a campus wide survey, individual interviews, and focus groups, targeting researchers from a wide range of disciplines. The findings from the survey provide valuable insight into the current RDM environment at the University of Victoria, with the potential to inform the planning and implementation of services and infrastructure to support researchers. In addition to gaining knowledge of researchers' own data management practices, this study also aimed to understand the role the University of Victoria Libraries could play.

The findings in this study highlight the diversity of data generated at the University of Victoria, which must factor in to planning decisions. A one-size fits all approach to RDM will only prove to create further challenges and barriers for researchers. The University of Victoria Libraries are well positioned to meet this challenge and encourage adoption of data management practices and tools throughout the research lifecycle, as they continue to develop their capacity for digitization and data curation, and develop workshops and training to meet the needs of digital scholars.

The following report recommendations are offered to guide the University of Victoria and the Libraries in responding to the wide range of challenges identified by the researchers surveyed in this study in managing and sharing their research data.

## 5.1 Recommendations for the University of Victoria Libraries

1. Develop discipline-specific workshops and training materials to help graduate students understand the importance of RDM.
2. Offer direct project-based support to help research teams to develop strong data management plans.
3. Develop workshops to help researchers document their data for reuse in other contexts.
4. Offer direct support at various stages in the research life cycle to tackle specific curation issues at the beginning, midway and after a research study.
5. Provide clear guidance on the distinction between active, archival, and repository storage and the role of each within the research data lifecycle.
6. Provide advice on repository options including discipline specific repositories, and repositories that are better suited for large data, or for particular data formats.
7. Offer guidance on journal policies and other scholarly communications requirements concerning RDM.
8. Promote the benefits of data sharing to university researchers, and help to remove some of the surmountable barriers to sharing.
9. Offer consultation on the retroactive sharing and curation of older data that may currently be at risk.
10. Work with faculty liaison librarians to determine their role in RDM support, and to better understand the specific needs of their departments and faculties.
11. Work in collaboration with RDM stakeholders across campus to improve communication channels, in order to effectively refer researchers to available supports and services.
12. Learn more about different community protocols, especially in the case of working with Indigenous data, to better address challenges and barriers to preservation.

## 5.2 Recommendations for the University of Victoria

1. Provide clear guidance on funder requirements concerning RDM.
2. Increase researcher awareness of institutional storage and backup options for working data.
3. Increase researcher awareness of Compute Canada default storage allocations, and assist researchers in gaining access to Compute Canada resources.
4. Provide discipline-specific guidance on standards for data description and formatting.
5. Identify RDM research champions at UVic to engage in RDM initiatives with the goal of expanding RDM capacity, expertise, and collaboration.
6. Increase the availability of sufficient, secure, easy to use storage solutions and RDM infrastructure to address current mandates and to meet future demand.
7. Continue to work nationally to advocate for increased funding for RDM infrastructure and expertise.

# Appendix A: Disciplinary Groupings

**Subject List**
Developed by the Canadian RDM Survey Consortium. Terms in bold are subject categories used to report findings. Bullet points are subjects that fall under each subject category.

**Engineering**
- Civil/mineral/mining/environmental engineering
- Biological/chemical/materials/mechanical engineering
- Electrical/computer engineering

**Science**
- Astronomy
- Biochemistry
- Biology
- Biophysics
- Cellular Biology
- Chemistry
- Computer Science
- Chemistry
- Earth Science
- Immunology
- Genetics
- Pharmacology
- Physics
- Mathematics
- Microbiology
- Molecular Biology
- Neuroscience
- Parasitology
- Pathology and Laboratory Medicine
- Toxicology

**Arts/Humanities**
- Art
- Art History and Conservation
- Asian Studies, Department of
- Classics
- Classics and Religious Studies
- Communication
- Cultural Studies
- Drama and Music
- English

- English Language and Literature
- Environmental Studies
- Film and Media
- Fine Arts
- French Studies
- Germanic and Slavic Studies
- History
- History/Classics
- Information Studies
- Jewish Studies
- Languages, Literatures and Cultures
- Library and Information Management
- Library, Archival and Information Studies
- Modern Languages
- Modern Languages and Literatures
- Music
- Official Languages and Bilingualism
- Philosophy
- Philosophy/Religion
- Religion
- Spanish
- Theatre
- Translation and Interpretation

**Business**

- Accounting and Finance
- Business
- Entrepreneurship and Innovation
- Finance
- Industrial Relations
- Management
- Management Information Systems
- Marketing and Behavioural Science
- Operations and Logistics
- Organizational Behaviour and Human Resources
- Strategy and Business Economics

**Education**

- Education

**Law**

- Law

**Social Sciences**

- Anthropology/Archaeology
- Criminology
- Economics
- Gender Studies
- Geography
- Geography / Planning
- Geography, Environment and Geomatics
- Global Development Studies
- International Development and Global Studies
- Linguistics
- Policy Studies
- Political Science
- Political Studies
- Psychology
- Public and International Affairs
- Public Administration
- Social Work
- Sociological and Anthropological Studies
- Sociology
- Urban and Regional Planning

**Health Sciences**

- Anatomy
- Anesthesia
- Dentistry
- Dermatology
- Geriatrics
- History of Medicine and Bioethics
- Kinesiology
    - Physical Education
    - Human Kinetics
- Medical Education
- Medicine (emergency, family, critical)
- Neurology
- Nursing and Midwifery
- Nutrition Sciences
- Obstetrics and Gynecology
- Oncology
- Ophthalmology
- Oto-rhino-laryngology
- Pediatrics
- Perioperative Medicine
- Physiology

- Population and Public Health
    - Epidemiology
    - Biostatistics
- Psychiatry
- Radiology and Diagnostic Imaging
    - Medical imaging
- Rehabilitation Sciences
    - Occupational Science and Occupational Therapy
    - Physical Medicine and Rehabilitation
    - Physical Therapy
    - Audiology and Speech-Language Pathology
    -
- Surgery
- Urology

**Interdisciplinary/Other**

- Departments not fitting into one of these groupings
- Not Specified

# APPENDIX B: INTEREST IN RDM SERVICE OFFERINGS BY RESPONDENT TYPE

**Table B1.** Interest ratings of respondents in research data management service offerings according to respondent type category (professor vs. graduate student).

| | PROFESSORS | | | | GRADUATE STUDENTS | | | |
|---|---|---|---|---|---|---|---|---|
| **SERVICE** | **Very Interested** | **Interested** | **Not Interested** | **Not Applicable** | **Very Interested** | **Interested** | **Not Interested** | **Not Applicable** |
| **WORKSHOPS ON BEST PRACTICES IN DATA MANAGEMENT FOR FACULTY** | 36 | 55 | 15 | 6 | 29 | 48 | 20 | 96 |
| **WORKSHOPS ON BEST PRACTICES IN DATA MANAGEMENT FOR GRADUATE STUDENTS** | 37 | 36 | 17 | 17 | 81 | 100 | 21 | 4 |
| **PERSONALIZED CONSULTATION ON DATA MANAGEMENT PRACTICES FOR SPECIFIC RESEARCH GROUPS OR PROJECTS** | 38 | 37 | 26 | 6 | 61 | 93 | 34 | 15 |
| **COMMUNICATION AND INFORMATION ABOUT FUNDING REQUIREMENTS AND JOURNAL REQUIREMENTS REGARDING RESEARCH DATA** | 31 | 55 | 12 | 11 | 62 | 101 | 27 | 10 |
| **ASSISTANCE PREPARING DATA MANAGEMENT PLANS TO MEET FUNDING REQUIREMENTS, OR ASSISTANCE CREATING FORMAL OR DOCUMENTED DATA MANAGEMENT POLICIES** | 42 | 44 | 15 | 8 | 64 | 103 | 27 | 7 |
| **DIGITIZATION OF PHYSICAL RECORDS** | 23 | 27 | 26 | 29 | 43 | 77 | 44 | 31 |
| **ASSISTANCE WITH DOCUMENTING AND DESCRIBING DATA (I.E. METADATA CREATION)** | 26 | 39 | 31 | 11 | 44 | 101 | 45 | 7 |
| **ASSISTANCE WITH ISSUES ASSOCIATED WITH DATA PRESERVATION AND/OR SHARING (CONFIDENTIALITY, PRIVACY, ETHICS, LEGAL, INTELLECTUAL PROPERTY RIGHTS)** | 35 | 49 | 15 | 8 | 61 | 96 | 38 | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **DATA STORAGE AND BACKUP DURING ACTIVE RESEARCH PROJECTS** | 40 | 36 | 28 | 5 | 71 | 87 | 32 | 7 |
| **AN INSTITUTIONAL REPOSITORY FOR LONG-TERM ACCESS AND PRESERVATION OF RESEARCH DATA** | 43 | 41 | 18 | 8 | 73 | 81 | 37 | 7 |
| **ASSISTANCE WITH DEPOSITING DATA IN APPROPRIATE DISCIPLINARY OR OTHER EXTERNAL DATA REPOSITORIES** | 17 | 52 | 27 | 10 | 36 | 93 | 56 | 13 |
| **ASSIGNMENT OF PERMANENT DIGITAL OBJECT IDENTIFIERS (DOIS) FOR DATASETS** | 20 | 46 | 24 | 16 | 38 | 81 | 58 | 20 |
| **ASSISTANCE IN FINDING AND ACCESSING DATA SOURCES** | 25 | 42 | 28 | 10 | 56 | 98 | 35 | 9 |

## APPENDIX C: SURVEY COVER LETTER

# Campus Wide Research Data Management (RDM) Survey at the University of Victoria

In order to become better prepared to support research data management (RDM), the University of Victoria Libraries, with the support of the Vice-President Research, are participating in a national study on the research data management needs and practices of faculty members, post-doctoral fellows, and graduate students. Your participation in this brief survey will help us to better understand how university researchers store, preserve, and disseminate research data from grant funded projects. This information will help prepare the University of Victoria to address the principles laid out in the Tri-Agency Statement of Principles on Digital Data Management, and will help the University of Victoria Libraries to design new services to support research data management. The aggregate data will feed into national initiatives lead by the Canadian Association of Research Libraries and the Leadership Council for Digital Infrastructure to develop infrastructure and services that will support better research data management for researchers across Canada.

The survey will ask you how you work with your research data, how you share your data, your awareness of funding mandates for research data management, and basic demographic questions about your professional status. There are only 15 short questions estimated time to complete the survey is approximately 10-15 minutes. The raw data will be electronically stored on secure servers maintained by the University of Victoria. Only members of the research team will have access to the raw data. Personally identifiable data will not be disclosed and will be stored only as is necessary for our research or to design services. Data will be stripped of contact and other identifying information, and will be shared only in aggregate format. Participants will not be named in any formal or informal products of the survey. Anonymized data will be deposited to a data repository and will be shared broadly. Destruction of the data may occur at a future time if and when the raw data do not provide any further research potential. If you would like a summary of the survey results, you may contact the Principal Investigator – Jacqueline Quinless (quinless@uvic.ca).

## APPENDIX D: RDM SURVEY QUESTIONS

**Section 1: Demographic & General Questions**

Q1. Please indicate your rank at UVic:

- ○ Graduate Student
- ○ Post-doctoral Fellow
- ○ Lecturer/Instructor
- ○ Librarian
- ○ Adjunct Professor
- ○ Assistant Professor
- ○ Associate Professor
- ○ Full Professor
- ○ Other, please specify... _____

Q2. Please indicate your discipline and home department._____

Q3. Please indicate your cross-discipline affiliations outside of your home department, if applicable._____

Q4. Which funding sources have you used within the past 5 years? Select all that apply:

- ☐ SSHRC Insight Grant
- ☐ SSHRC Partnership Grant
- ☐ SSHRC other, please specify _____
- ☐ Canada Council for the Arts
- ☐ CIHR
- ☐ CFI
- ☐ NSERC
- ☐ ARC (UK)

- ☐ ESRC (UK)
- ☐ EU
- ☐ Industry
- ☐ Mellon Foundation
- ☐ MITACs
- ☐ NEH (USA)
- ☐ NIH (USA)
- ☐ SSHRC (USA)
- ☐ None
- ☐ Other, please specify... _____

**Section 2: Working with Research Data**

In this section you will be asked questions about your research data, including how you work with them, document them and store them

Q5. Which of the following best describes the type of research data you generate or use in a typical research project? Select all that apply:

- ☐ Geospatial - (e.g. raster, vector, grid)
- ☐ Instrument specific – (e.g. Olympus Confocal Microscope Data Format, FLIR Infrared Camera (SEQ))
- ☐ Models – (e.g. 3D, statistical, similitude, macroeconomic, causal)
- ☐ Multimedia– (e.g. JPEG, TIFF, MPEG, MP3, Quicktime, Bitmap)
- ☐ Numerical – (e.g. CSV, MAT, XLS, SPSS)
- ☐ Software– (e.g. Java, C, Perl, Python, Ruby, PHP, R)
- ☐ Text - (e.g. TXT, DOC, PDF, RTF, HTML, XML)
- ☐ Other (e.g. discipline specific such as CIF, FITS, DICOM) please specify: _____

Q6. Please list any software and/or hardware used for the collection, analysis, or manipulation of your research data, if applicable:

Q7. Please indicate where you store research data from your current project(s). Select all that apply

- ☐ Flash drive/USB
- ☐ CD/DVD
- ☐ Computer hard drive (i.e. local hard drive)
- ☐ Laptop hard drive
- ☐ External hard drive
- ☐ Hard drive of the instrument/sensor which generates the data
- ☐ Shared drive/university or departmental server
- ☐ Cloud/web based solution (e.g. Dropbox, Google Drive, Amazon Cloud, Microsoft Cloud)
- ☐ External data repository (e.g. Institutional Repository, GitHub, tDAR, CWRC, Figshare, HathiTrust)
- ☐ Grid/high performance computing (HPC) centre
- ☐ Physical copy retained (in boxes, cabinets, etc.)
- ☐ Not sure
- ☐ Other, please specify: _____

Q8. Do you think there is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person that is part of your research team to understand and use the research data?

- ○ Yes
- ○ No
- ○ Not sure

Q9. Do you think there is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person that is NOT part of your research team to understand and use the research data?

- ○ Yes
- ○ No
- ○ Not sure

If no, please explain:_____

Q10. Which methods of sharing your research data do you currently use? Select all that apply. If you do not currently share your data, choose 'not currently sharing'.

- ☐ Not planning to share
- ☐ Share by personal request
- ☐ Share online with restricted access
- ☐ Upload online to an institutional or personal website
- ☐ Upload online to an institutional repository, such as Dataverse
- ☐ Include as part of supplementary material files to a journal publisher
- ☐ Deposit in a general or discipline-specific repository, such as GitHub, tDAR, CWRC, Figshare, HathiTrust. Please specify: _____
- ☐ If you plan to use another data repository other than the ones listed above, please specify: _____

Q11. Some research data cannot be shared because of legal or privacy restrictions or embargoes. Which of the following restrictions or embargoes may limit your ability to share your data with others? Select all that apply. If there are no restrictions or embargoes, choose 'there are no restrictions or embargoes on sharing my data with other parties'

- ☐ There are no restrictions or embargoes on sharing my data with other parties
- ☐ I need to publish my data before I can share them

- ☐ Sharing my data may jeopardize Intellectual Property rights
- ☐ I plan to file for a patent
- ☐ My data cannot be shared because of commercial concerns
- ☐ I have a contractual obligation with a third party
- ☐ My data are subject to privacy, confidentiality, or ethics restrictions (e.g. survey data with personal information)
- ☐ My data are a matter of public safety or of a sensitive nature
- ☐ I'm unsure if I am allowed to share my data
- ☐ Other, please specify: _____

Q12. What, if any, are the reasons you would not be willing to share your research data and associated methods/tools? Select all that apply. If you are willing to share, choose 'I am willing to share them'.

- ☐ They are incomplete or not finished
- ☐ I still wish to derive value from them
- ☐ I do not have the technical skills or knowledge
- ☐ I do not hold the rights to share them
- ☐ Funding body does not require sharing
- ☐ I believe they should not be shared
- ☐ I did not know I could share them
- ☐ Insufficient time
- ☐ Lack of standards to make them usable by others
- ☐ Lack of funding
- ☐ No place to put them
- ☐ They are not useful to others
- ☐ There are privacy, legal or security issues
- ☐ My data could potentially be used without proper citation
- ☐ I'm concerned my data could be used without proper citation or acknowledgement
- ☐ I am willing to share them
- ☐ Other, please specify: _____

Q13. What benefits do you see to sharing your research data? Select all that apply. If you see no benefits, choose 'I see no benefits to sharing my data

- ☐ I see no benefits to sharing my data
- ☐ Data availability provides safeguards against misconduct, data fabrication and falsification
- ☐ Data sharing and/or replication studies help in the training of next generation researchers
- ☐ Data sharing enables my data to be cited and increases my research impact
- ☐ Data sharing encourages collaborative scholarship
- ☐ Data sharing encourages interdisciplinary research
- ☐ Data sharing moves my field of research forward
- ☐ Data sharing reduces redundant data collection
- ☐ Data sharing supports open access to knowledge
- ☐ Re-analysis of data helps verify results
- ☐ Well-maintained data helps retain data integrity
- ☐ Other, please specify: _____

Q14. Data management plans typically address questions about research data types and formats: standards to be used for describing data; ethics and legal compliance; plans for preservation, access, sharing, and reuse; and responsibilities assigned and resources needed. If you were asked to draft a data management plan as part of a grant application, which of the following statements would best describe your situation? Select one:

- ○ I would be able to draft a data management plan that would address these types of questions without assistance
- ○ I would be able to draft a data management plan that would address these types of questions, but would prefer to have assistance and/or guided documentation to ensure the success of my application
- ○ I would need assistance and / or guided documentation to appropriately address some or all of the sections

Q15. If data management plans were made part of grant applications from funding bodies such as SSHRC, CIHR, and NSERC, how interested would you be in the following services? Please rate your interest in each service. If the service does not apply to your situation, choose 'not applicable'.

| | very interested | interested | not interested | not applicable |
|---|---|---|---|---|
| Workshops on best practices in data management for faculty. | ○ | ○ | ○ | ○ |
| Workshops on best practices in data management for graduate students. | ○ | ○ | ○ | ○ |
| Personalized consultation on data management practices for specific research groups or projects. | ○ | ○ | ○ | ○ |
| Communication and information about funding requirements and journal requirements regarding research data. | ○ | ○ | ○ | ○ |
| Assistance preparing data management plans to meet funding requirements, or assistance creating formal or documented data management policies. | ○ | ○ | ○ | ○ |
| Digitization of physical records | ○ | ○ | ○ | ○ |
| Assistance with documenting and describing data (i.e. metadata creation). | ○ | ○ | ○ | ○ |
| Assistance with issues associated with data preservation and/or sharing (confidentiality, privacy, ethics, legal, intellectual property rights). | ○ | ○ | ○ | ○ |
| Data storage and backup during active research projects. | ○ | ○ | ○ | ○ |
| An institutional repository for long-term access and preservation of research data. | ○ | ○ | ○ | ○ |
| Assistance with depositing research data in appropriate disciplinary or other external data repositories. | ○ | ○ | ○ | ○ |
| Assignment of permanent digital object identifiers (DOIs) for datasets. | ○ | ○ | ○ | ○ |
| Assistance in finding and accessing data sources. | ○ | ○ | ○ | ○ |

## REFERENCES


Government of Canada (2014). Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation. http://www.ic.gc.ca/eic/site/113.nsf/eng/h_07657.html

Government of Canada (2016). Tri-Agency Statement of Principles on Digital Data Management. https://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html

Government of Canada (2018). Draft Tri-Agency Research Data Management Policy for Consultation. https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html

Panel on Research Ethics (2008). Draft 2nd Edition of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. http://www.pre.ethics.gc.ca/policy-politique/docs/TCPS-Draft2-eng.pdf

Portage (2018a). About Portage. Accessed November 08, 2018. https://portagenetwork.ca/about/

Portage (2018b). Canadian RDM Survey Consortium. Accessed November 08, 2018. https://portagenetwork.ca/working-with-portage/network-of-expertise/rdm-survey-consortium

University of Victoria (2018). Annotated Guidelines for Completing the Human Research Ethics Board Application for Ethics Approval for Human Participation Research. https://www.uvic.ca/research/assets/docs/Ethics/Annotated%20Guidelines%20HRE%20Application%20Form.pdf


## FURTHER READING


Barsky, E., Mitchell, M., Buhler, J. (2016). UBC Research Data Management Survey. http://hdl.handle.net/11272/10364. V1.

Hwang, C. (2016). Research Data Management Faculty Survey, University of Alberta. https://drive.google.com/file/d/0B0Y3YAsi7aKpdkxzMkVfVDdadzg/view

Irwin, S. (2013). Qualitative Secondary Data Analysis: Ethics, Epistemology, and Context. *Progress in Development Studies*, 13(4), 295-306. doi: 10.1177/1464993413490479

Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Eds.). (2014). Privacy, Big Data, and the Public Good: Frameworks for Engagement. New York, NY: Cambridge University Press.

Metcalf, J., & Crawford, K. (2016). Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. *Big Data and Society*, 3(1). doi: 10.1177/2053951716650211

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57(6), 1701-1778. http://www.uclalawreview.org/pdf/57-6-3.pdf

Ohm, P. (2015). Sensitive Information. *Southern California Law Review*, 88(5), 1125-1196. https://southerncalifornialawreview.com/2015/07/02/sensitive-information-article-by-paul-ohm

Sewerin, C., Dearborn, D., Henshilwood, A., Spence, M., & Zahradnik, T. (2015). Research Data Management Faculty & Post-Doctoral Survey, University of Toronto. http://hdl.handle.net/1807/69144

Social Sciences and Humanities Research Council (2013). Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada. http://www.sshrccrsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf

Sweeney, L., Crosas, M., & Bar-Sinai, M. (2015). Sharing Sensitive Data with Confidence: The DataTags System. *Journal of Technology Science*. https://techscience.org/a/2015101601

Szigeti, K. (2016). University of Waterloo Survey Instrument Dec. 2015. V2 [Version]. http://hdl.handle.net/10864/11649.

Zaraiskaya, T., Cooper, A., Moon, J., Murphy, S., Saleh, N. (2016). Queen's University RDM Survey. Data Services V1. http://hdl.handle.net/10864/11651.